

# G

---

## Galerkin Method

- [Finite Element Method for Fluid Film Bearings](#)

---

## Galling

- [Gear Contact Temperature and Scuffing Risk Analysis](#)

---

## Gas Bearing Applications

CHRISTOPHER DELLA CORTE

Tribology and Surface Science, NASA Glenn Research Center, Cleveland, OH, USA

### Synonyms

[Gas lubrication applications](#); [Hydrodynamic gas bearing applications](#); [Oil-free bearing applications](#); [Oil-less bearing applications](#)

### Key Applications

Air cycle machines, oil-free blowers and compressors, gyroscopes, gas circulators, turbochargers, turbines

### Definition

Rotating shaft bearing applications that utilize gas-lubricated bearings for rotor support typically do so to avoid oil-related contamination issues, to operate at extraordinary high speeds and temperatures, or, in the case of hydrostatic bearings, to achieve very low friction.

### Scientific Fundamentals

#### Underlying Mechanism

In a bearing, effective lubrication occurs when an easily sheared intermediate layer or film separates moving and stationary surfaces. Deformation or velocity

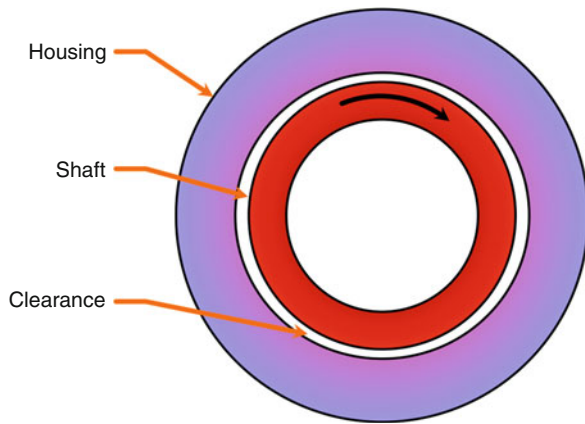
accommodation occurs within this film, resulting in low friction and wear. When a gas film under pressure is used for this separation function, a gas bearing exists. Gas pressure can be intentionally supplied from external sources, as in hydrostatic gas bearings, or can develop through the self-acting hydrodynamic effect.

### Bearing Types and Classifications

Gas bearings can be self-acting, hydrostatic, or employ a combination or hybrid design to provide the pressure required to support design loads. In addition, bearing geometry can be rigid, with fixed clearances, utilize rigid pads that are flexibly supported, or be made from compliant materials, typically sheet metal foils. These bearing design variations, discussed in more detail elsewhere in this encyclopedia, often impact the type of gas bearing applications in which discrete bearing types are found. [Figure 1](#) shows the cross section diagram of the most basic type of gas bearing: the self-acting, fixed clearance journal bearing (Constantinescu et al. 1985). [Figure 2](#) is a photograph of such a bearing.

### General Gas Bearing Applications

Many engineering and scientific machine applications involve bearings that must provide shaft load support and dynamic stability without limiting operating speed and temperature or being a source of contamination often associated with conventional oil-lubricated rotor support systems. In such specialty applications where oil cannot be used, gas bearings may be employed. Applications that can be characterized by steady-state environmental conditions (speed, temperature, little or no dynamic-shock loads) can often be satisfied with rigid-type self-acting or hydrostatic gas bearings. Examples of such applications include high-speed machine tool spindles, high-speed gyroscopes, and turbo-compressors for small stationary cryo-coolers. Applications in which the speed and temperature ranges greatly, or in which high levels of damping, shock resistance, load capacity and tolerance to misalignment, shaft distortions, and debris resistance are required are satisfied with compliant surface gas bearings, commonly referred to as foil bearings (Barnett and Silver 1970). More specific details on rigid



**Gas Bearing Applications, Fig. 1** Self-acting rigid journal gas bearing. Pressure is developed through the hydrodynamic effect. The clearance is fixed and is typically 5–25  $\mu\text{m}$  and the eccentricity varies according to load, speed, and design



**Gas Bearing Applications, Fig. 2** Photograph of simple rigid gas journal bearing shaft (front) and sleeve (rear)

style gas bearings and compliant surface type bearings are found in other entries of this encyclopedia.

### Rigid Gas Bearing Applications

Few rotating machines can be supported on rigid gas bearings. Precisely controlled dimensions, operating clearances, alignment, and environmental factors typify these applications. The fluid film pressure can be generated in

one of two ways, external pressurization or through the hydrodynamic effect. For pressurized bearings, often called hydrostatic, the science and technology is well understood and well developed (Gross 1962). Readers are referred to the literature for design and performance guidelines. In general, the operating pressure is integrated over the bearing area to estimate the load capacity. Stiffness levels are high and stability can be achieved through careful orifice and pocket design details.

Such hydrostatic bearings are in common use as machine tool spindle bearings. In this application, precise control of the spindle position and orbit can be achieved and thus superior finishes can be obtained. If a single-point diamond tool is used with a hydrostatically supported spindle in a turning lathe, optical quality mirror surfaces are routinely manufactured. Further, near zero friction and very high speeds can be obtained with ease. The downside to this approach is the need to develop and maintain extraordinary precision components and the energy needed to provide adequate supply air pressure and cleanliness.

For rigid geometry bearings that are self-acting hydrodynamic in design, market penetration of such devices has been negligible. In the past, several research machines have been built, mostly in the fields of nuclear reactor gaseous coolant pumps, but these were never commercialized. High-speed gyroscopes, using ceramic journal and thrust bearings lubricated with pure gases such as hydrogen, have been deployed in exotic space systems but more modern approaches and solid-state devices have replaced these. Such rigid, hydrodynamic bearings develop low gas pressures and thus have low load capacities except at exceedingly low film thicknesses and high speeds. High speeds result in heat generation—derived thermal expansion and thin gas films lead to the potential for rubbing contact, wear, debris generation, and bearing failure. Further, both self-acting and hydrostatic rigid bearings suffer from a lack of damping often needed to quell rotor vibration and instabilities. Damping in rigid bearings must come from the gas films that have such low viscosities and thus provide little useful damping.

A more recent effort to develop improved, non-circular geometry rigid bearings using modern design and manufacturing methods has confirmed the limited utility of such devices. Load capacity is generally low and, while such bearings can be made intrinsically stable, they provide no appreciable shaft damping capability. These lobed bearings, however, when lubricated with a viscous fluid such as oil using increased clearances, have been shown to be suitable for more conventional oil-lubricated applications.

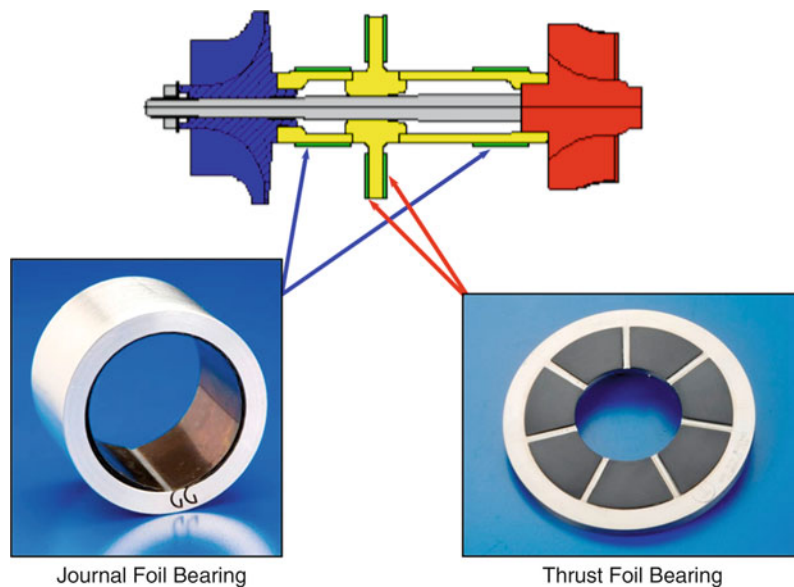
### Compliant Surface Bearing Applications

Compliant surface gas bearings, often called foil gas bearings, are a direct descendent of rigid geometry gas bearings. Foil bearings are described in detail in another section but their basic premise is reiterated here for clarity. Foil bearings operate under the same fluid mechanics principles as rigid geometry hydrodynamic bearings expect that the geometry of the bearing element is not fixed. In fact, in foil bearings, the foil structural stiffness is sufficiently compliant to allow the modest hydrodynamic gas film pressure to deform the foil significantly. In this manner, a thicker, more uniform gas film results and this increases bearing load capacity. In addition to providing a thicker gas film, the compliance of the foils and their support structure lead to useful ancillary properties such as tolerance to debris and misalignment, thermal distortions and shaft expansion, and distortions that typically accompany high-speed rotating machinery. When combined with high- or low-temperature solid lubricants that prevent excessive friction and wear during start up and shut down operation, foil bearings enable a variety of fully commercialized and emerging oil-free rotor support applications (Klaass et al. 2006). Figure 3 provides photos of journal and thrust foil bearings and their locations on a developmental oil-free turbocharger rotor utilizing a bearing layout often called “straddle mounted,” which is typical for foil bearing-supported machines.

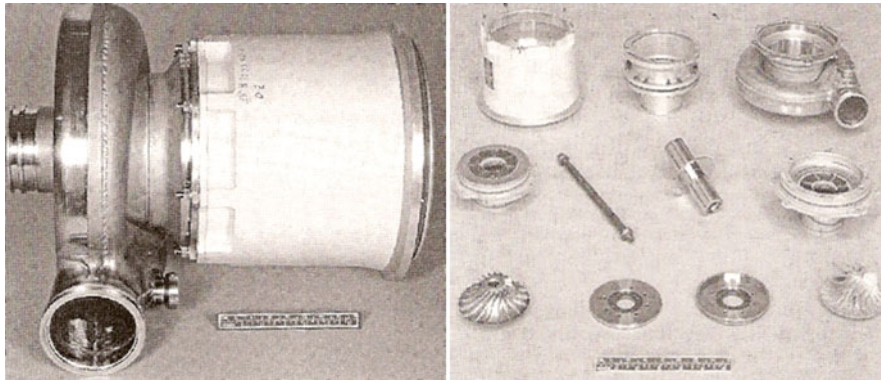
The first and most prevalent applications for compliant surface (foil) gas bearings are in the air cycle machine (ACM) compressors used for aircraft environmental control systems. These machines use engine bleed air to drive turbine shafts that rotate turbocompressors to pressurize and cool aircraft cabin air. The ability to preclude oil contamination is the major driver for the use of oil-free foil gas bearings in this application. Today, all commercial aircraft utilize foil bearing supported ACMs, and these have proven to be very reliable and robust, offering more than 100,000 h mean time between failures (mtbf). Figure 4 shows a typical ACM supported on gas foil bearings.

Following this success in ACMs, foil bearings were adopted into the cryogenic fluid handling industry in the form of turbocompressor and turboexpander bearings. Again, the advantage of such bearings lies in the combination of prevention of contamination and ability to operate at extremely low temperatures and high speeds. Both major equipment manufacturers and smaller entities produce machines for these applications.

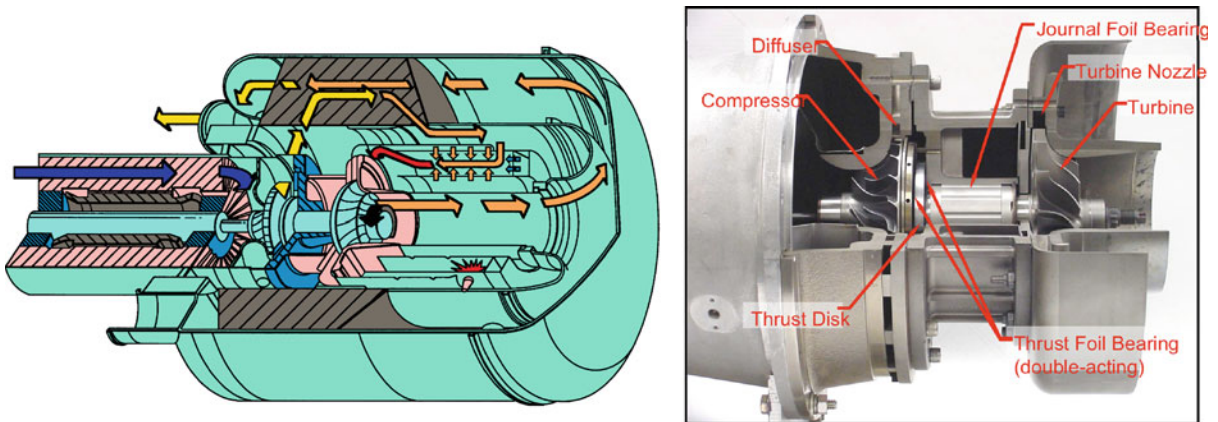
Compliant surface gas bearings are also useful for high-speed, high-temperature applications such as turbochargers and small gas turbine engines. While several manufacturers have demonstrated oil-free microturbine engines utilizing foil bearings, only the Capstone Turbine Corporation has a fully commercialized product line.



**Gas Bearing Applications, Fig. 3** Oil-free turbocharger cross section showing typical layout of thrust and journal bearings used for load support



**Gas Bearing Applications, Fig. 4** Early air cycle machine application of foil bearings showing the assembled machine (*left*) and the individual components (*right*)



**Gas Bearing Applications, Fig. 5** Schematic diagram (*left*) and cutaway photograph (*right*) of 30 kW Capstone Turbine utilizing foil air bearings

Their machines include foil bearings in the engines and generators and are available in sizes ranging from 30 to 200 kW electrical output. Figure 5 shows a sketch and photo of a Capstone 30 kW turbine and Fig. 6 shows the packaged system installed at the NASA lab.

Recently commercialized industrial gas compressors and blowers capitalize on foil as bearings for lowering maintenance costs and improving efficiency through lowered friction. Such turbocompressors are available from 30 to 400 kW and provide oil-free process air from low pressure (blowers) that operate at a few atmospheres discharge pressure to multistage machines that provide pressure ratios of over 10.0. These machines are typically driven by electric motors mounted directly on the rotating

shaft that also drives centrifugal or axial compressor wheels. Such machines are noted for their simplicity and reliability and are predominantly produced in Korea.

Emerging applications for compliant surface gas bearings include oil-free turbochargers and hybrid shaft support systems in which more than one type of bearing is used for rotor support. Major turbocharger manufacturers have demonstrated, and in some cases patented, oil-free turbochargers for internal combustion engine applications. It is expected that these machines will be marketed in the near future. Other likely high-production-volume applications are the blowers and compressors used for fuel cell air management. In this case, clean, compressed air supply without any possibility of oil





**Gas Bearing Applications, Fig. 6** Capstone C30 oil-free microturbine package installed at NASA labs

contamination, which would poison a fuel cell stack, is the driving force behind this and many other gas bearing supported machines.

Gas bearings provide generally lower load capability than conventional bearing technologies. But in applications in which high speeds, extreme temperatures, and sensitivity to lubricant related contamination abound, gas bearings can be a very attractive solution.

## Cross-References

- [Cryogenic Hybrid Fluid Film Bearings](#)
- [Foil Gas Bearings](#)
- [Hydrostatic/Hybrid Gas Bearings](#)
- [Tilting Pad Gas Bearings](#)

## References

- M.A. Barnett, A. Silver, *Application of Air Bearings to High-Speed Turbomachinery*. SAE Paper 700720, 1970
- V.N. Constantinescu, A. Nica, M.D. Pascovici, G. Ceptureanu, S. Nedelu, *Sliding Bearings* (Allerton Press, New York, 1985)
- W.A. Gross, *Gas Film Lubrication* (Wiley, New York, 1962)
- R.F. Klaass, C. DellaCorte, The quest for oil-free gas turbine engines. In *Proceedings of the Power Systems Conference*, New Orleans, LA, SAE Paper 2006-01-3055, Nov 2006

## Gas Bearing Materials

CHRISTOPHER DELLA CORTE

Tribology and Surface Science, NASA Glenn  
Research Center, Cleveland, OH, USA

## Synonyms

[Bearing materials](#); [Coatings as gas bearing materials](#); [Solid lubricants for gas bearings](#); [Wear-resistant materials for gas bearings](#)

## Key Applications

Air cycle machines, oil-free blowers and compressors, turbochargers, turbines

## Definition

Gas bearings are comprised of a rigid moving surface and a stationary surface that may be rigid or, in the case of compliant surface gas bearings, elastic and flexible. The materials used for gas bearings must provide smooth surfaces that can withstand occasional dry rubbing contact and remain smooth to enable the formation of thin gas films. Low friction and wear, often at extreme temperatures, are additional characteristics of gas bearing materials.

## Scientific Fundamentals

### Rigid Gas Bearing Materials

Materials employed for rigid gas bearings can be divided into two broad categories, structural materials and tribological materials, and as such provide distinct but sometimes overlapping functions. The underlying key function of gas bearing materials is to provide smooth shaft and bearing surfaces that allow complete physical separation of the moving from non-moving members utilizing a gas film only a few microns thick. In addition to being able to be manufactured to low roughness and high precision, gas bearing materials must simultaneously remain in a smooth condition even after repeated start-stop rubbing cycles and exposure to debris and possibly contamination. Lastly, given the precision levels needed to establish and maintain thin gas films, low friction surfaces must be provided to minimize thermal effects that ensue from intermittent rubbing contact (Murray 1969).

Rigid gas bearings, described in more detail in other sections, are typically made from steels, bronzes, and ceramics. All of these structural materials can be manufactured with high precision and smooth surface finishes. Further, these materials can typically withstand

limited sliding contact without galling or otherwise becoming roughened by the wear process. For steel shafting, electroplated coatings of nickel or chrome are often employed to further enhance corrosion resistance and durability. More recently, thin ceramic coatings, like titanium nitride and titanium carbide, have been used over stainless steel with some success. For gyroscope bearings, monolithic alumina has been employed because of its excellent dimensional stability, mirror-smooth finished surface capability, and low thermal expansion coefficient that enables the maintenance of small clearances over a wider temperature range.

In the case the rigid bearings, the counterpart to the shaft surfaces, the bearings themselves, are typically made from bronze, graphite, or other free-running, non-galling tribological material. For hydrostatic bearings, holes, passageways, orifices, and recessed pockets are machined into the bearings depending upon their design. Thus, easily machined materials are preferred over hard ceramic materials. For self-acting rigid gas bearings, monolithic ceramics have been demonstrated, at least at the laboratory scale though the cost is undoubtedly high. In some instances, such as tilting pad gas bearings, pads have been made from metal-bonded titanium carbide. These composite materials are hard, dimensionally stable, and can be made through near net shape powder metallurgy techniques to minimize costs and achieve smooth finishes and precise final geometries.

Bearings made from graphite require no additional solid lubrication. Bearings made from ceramics and other metals are often coated with low-friction graphite or diamond-like carbon films, contain regions of solid lubricant to act as a lubricant transfer source, or employ hydrostatic pressure before motion is initiated to prevent friction and wear. Whatever materials are chosen, key attributes that must be retained are precise geometries and dimensions, smooth surface finishes, and low friction. [Figure 1](#) shows a photograph of a rigid gas bearing with the smooth, highly precise surfaces and manufacturing qualities typical for such bearings.

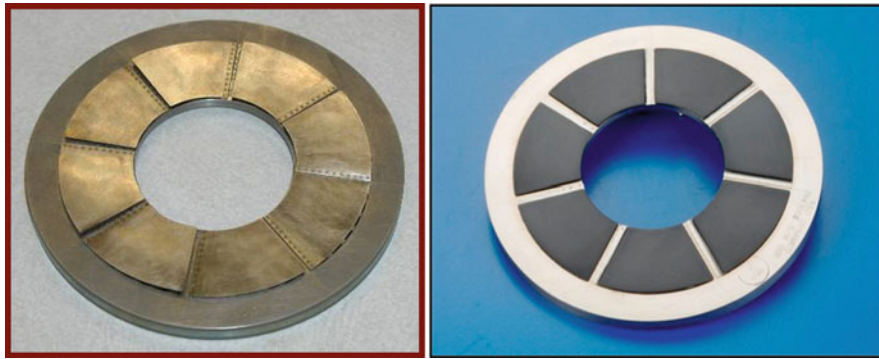
### Compliant Surface Foil Bearings

Foil bearings are a subset of gas bearings in which a rigid shaft operates against a smooth flexible sheet metal foil surface. The foil surface floats on a gas film during normal operation but rubbing contact occurs during start-up and shutdown and whenever loads exceed the film pressure. Just like in rigid bearings, the materials must be carefully chosen to maintain smooth surfaces, minimize friction, and withstand environmental temperature extremes and debris ingestion.



**Gas Bearing Materials, Fig. 1** Rigid gas bearing employing hard, polished shaft and sleeve surfaces to achieve high levels of precision required for operation on thin gas films. Surface coatings employed, such as chrome carbide or thin dense chrome, prevent stiction, wear, and galling

Shaft or runner materials are typically steels or stainless steels that have been coated with nickel or chrome to enhance wear resistance, prevent corrosion, and mitigate galling tendencies. These shaft treatments operate against sheet metal foils that are typically only 0.2 mm thick and made from nickel-based superalloys like Inconel X-750. Historically, other alloys such as beryllium-copper have been used, but the nickel-based superalloys are almost exclusively used today for a variety of reasons. They have excellent spring properties from cryogenic temperatures to over 600°C. They can be easily formed in their annealed state and then heat treated to provide necessary final properties. They are fatigue resistant, corrosion “proof,” and compatible with many common polymer solid lubricant coatings. Lastly, they are readily available as smooth foils from a variety of sources. To retain a smooth finish and lower friction, the nickel-base superalloys do require solid lubrication. This can come from soft surface coatings like PTFE and bonded graphite or from transfer of solid lubricants from a shaft coating (Bhushan et al. 1980). [Figure 2](#) shows a foil thrust bearing before and after coating with a polyimide coating doped with solid lubricant pigments. [Figure 3](#) shows foil bearing wear surfaces after repeated start stop cycles and operation at high speeds and loads. Wear patterns emerge that



**Gas Bearing Materials, Fig. 2** Thrust foil bearing uncoated (*left*) and coated with a soft polymer solid lubricant coating. The foil coating reduces sliding friction and enables a rapid “break-in” of the bearing in which surface irregularities and “high spots” wear away quickly, leading to a conformal bearing-mating surface topography

G



**Gas Bearing Materials, Fig. 3** Worn foil bearing surfaces showing polishing wear “witness marks” where foil surface has touched shaft or runner during start-up and shut down. Thrust bearing surface (*left*) and journal foil bearing (*right*) surface wear, often called “tiger stripes,” where the wear occurs preferentially in areas directly supported by bump foil peaks

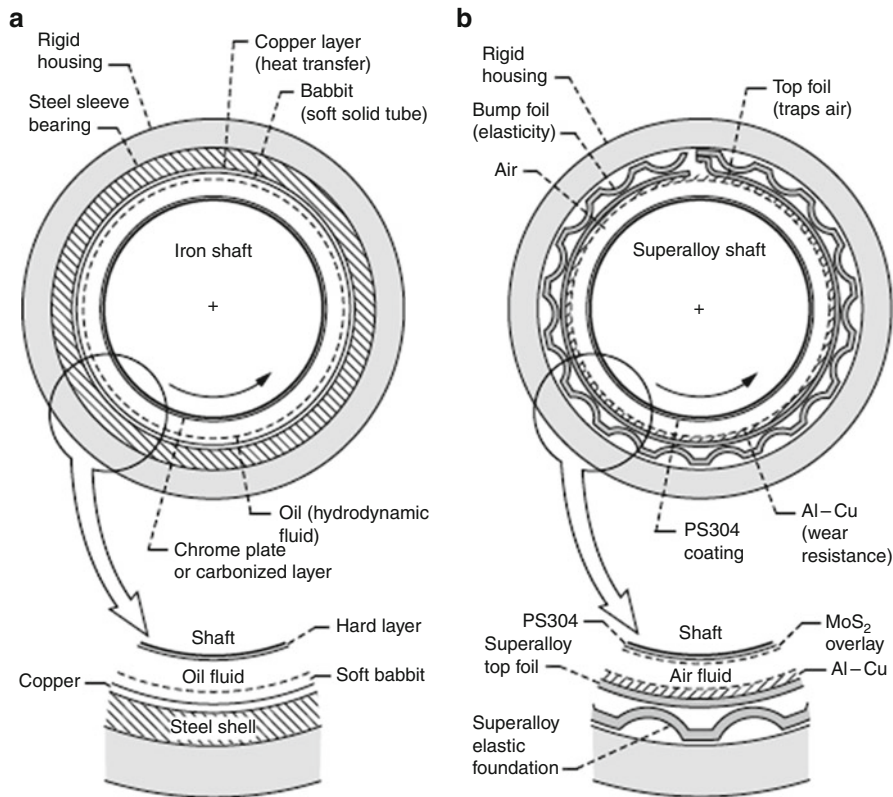
appear as stripes or bands that coincide with the peaks of underlying bump foil structural members.

The most prevalent shaft coatings for operating against foil bearings are chrome-plated steel and chrome-plated nickel-based superalloys. Nickel coatings are also used. These hard platings are ground and polished to enhance the surface finish and prevent galling should the foil coatings wear through to the underlying foil substrate in a localized area. For bearings that are intended for long life at extremely high temperatures, several emerging composite shaft coatings and foil coatings tailored with high temperature solid lubricants are used.

The NASA-developed PS200 and PS300 series of shaft coatings employ the composite approach to foil bearing lubrication. These coatings combine wear-resistant, thermochemically stable metal bonded ceramic matrices

with solid lubricant additions to provide moderate friction and wear over a wide temperature range. Silver is used as a low temperature lubricant and a fluoride eutectic is used as the high temperature lubricant. In use, a solid lubricant glaze forms on the coated surface and a transfer film of solid lubricants forms on the foil mitigating friction and wear. A break-in period at high temperature or a sacrificial overlay of solid lubricant, like molybdenum disulfide, must be used to achieve maximum bearing performance. Bench testing of bearings and engine tests confirm the use of such coatings in foil bearings (Lubell et al. 2006).

Alternatively, for lower life applications, several commercially available solid lubricant pigmented foil coatings can be used against thin, dense, chrome-plated shafts. Such coatings typically contain molybdenum and tungsten disulfide, graphite, boron nitride, and other materials



**Gas Bearing Materials, Fig. 4** Solid lubricant system for (a) oil-lubricated bearings is similar to (b) foil gas bearings

co-deposited with silica or nickel-based bonding agents. Multiple layers and heat treatment steps are required to build up a relatively thin layer on the foil surface ( $\sim 10 \mu\text{m}$  thick). The use of such foil coatings offers two primary advantages. One is that low-friction traditional solid lubricants can be used, which provides reduced sliding torque at the onset of bearing use. That is, no break-in period is required. The second advantage is that any errors in bearing geometry resulting from poor design or manufacturing that result in “high spots” are rapidly worn away, leading to an improved bearing geometry in a manner similar to the abradable seal coatings used in turbine engines. In any case, such coatings can be used in concert with even the NASA coatings to give overall acceptable performance (Heshmat et al. 2005; Jahanmir et al. 2009).

Gas bearing materials are called upon to provide many characteristics. Each application requires a careful assessment for proper materials selection. In the end, the material selection is an integral part of the bearing system in much the same way as lead tin babbitt in an oil-lubricated bearing is part of a conventional bearing “system,” as shown in Fig. 4. Depending upon the life, temperature,

and other factors inherent to a specific machine, the bearing type, construction materials, and tribological materials selected will differ appropriately.

## Cross-References

- [Bonded Solid Lubrication Coatings, Process and Applications](#)
- [Cryogenic Solid Lubrication](#)
- [Fluid Film Bearing Materials](#)
- [Foil Gas Bearings](#)
- [Gas Bearing Applications](#)
- [Hydrostatic/Hybrid Gas Bearings](#)
- [Solid Lubricants](#)
- [Tilting Pad Gas Bearings](#)

## References

- B. Bhushan, D. Ruscitto, S. Grey, *Development of a Hydrodynamic Air Lubricated Surface Bearing for an Automotive Gas Turbine Engine-Part II-Materials and Coatings*. NASA CR-159848, 1980
- H. Heshmat, P. Hryniewicz, J.F. Walton II, J.P. Willis, S. Jahanmir, C. DellaCorte, Low-friction wear-resistant coatings for high-temperature foil bearings. *Tribol. Int.* **38**, 1059–1075 (2005)
- S. Jahanmir, H. Heshmat, C. Heshmat, Assessment of tribological coatings for foil bearing applications. *Tribol. Trans.* **52**(2), 231–242 (2009)



D. Lubell, C. DellaCorte, M.K. Stanford, Test evolution and oil-free engine experience of a high temperature foil air bearing coating In *Proceedings of GT2006:ASME Turbo Expo 2006*, Barcelona, Spain, 8–11 May 2006. Paper number GT2006-90572

S.F. Murray, *Research and Development of High Temperature Gas Bearings*. NASA CR-MTI-CR-22, 1969

## Gas Bearings

### ► Reynolds Equation for Compressible Fluid or Gas Film

## Gas Bearings with Narrow Inclined Grooves

CODA H. PAN

Global Technology, Millbury, MA, USA

### Synonyms

Chevron patterned journal bearing; Helical grooved journal bearing; Herringbone grooved journal bearing; Spiral groove bearings; Spiral grooved conical bearing; Spiral grooved opposed-hemispherical bearing; Textured gas bearings; Whipple thrust bearing

### Definition

Texturing of narrow inclined grooves on either surface of a rotating bearing induces pressurization and flow in the direction normal to the circumferential surface velocity.

### Scientific Fundamentals

An idealized cross-section of the grooving texture is rectangular, as shown in Fig. 1. The texture step acts to create

an abrupt change of the normal pressure gradient that induces a transverse flow (averaged across a period of the texture pattern) along  $\vec{j}$  that is paired with a corresponding pressure gradient in the manner of the pressure-flow characteristics of a pump. A particular bearing design is analogous to placing a viscous pump into a service loop.

### Types of Textured Gas Bearings

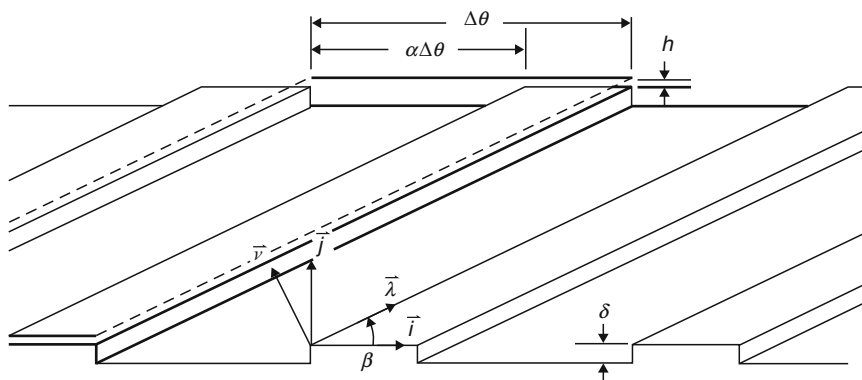
Spiral groove texturing was originally conceived for enabling thrust support with alternative plan forms shown in Fig. 2.

The concept has been extended to the cylindrical journal bearing to achieve high stiffness and enhanced dynamic stability. Figure 3 illustrates a plan form that features oppositely inclined helical grooves separated by a smooth band.

In the cartridge form, disk and cylindrical surfaces with inclined narrow groove texturing can be combined for omni-directional load support. Or, as shown in Fig. 4, opposed conical frustums can be paired to form an omni-directional support capability with a very high angular stiffness. Opposed hemispheres can be used in place of conical frustums, as shown in Fig. 5.

A sphere, in the form of a ball attached to a shaft, see Fig. 6, can also be textured with narrow spiral grooves; such a bearing controls linear displacement of the center of the supported rotor in all three directions of a Cartesian coordinate system while allowing free motion in all angular degrees of freedom to the extent that the through shaft does not interfere with the holes of the rotor.

Each of the various plan forms can be characterized by the radius profile of the bearing.



Gas Bearings with Narrow Inclined Grooves, Fig. 1 Bearing surface with texture of narrow inclined grooves

Analysis/Design of Inclined-Groove Textured Bearings

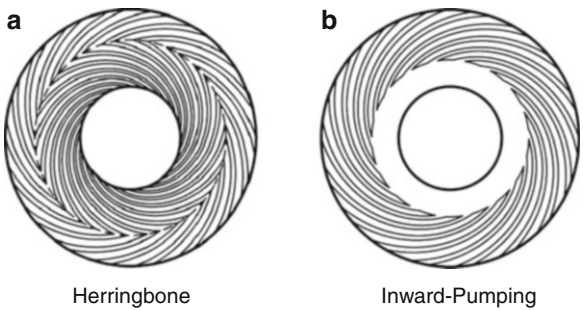
A study of an inclined-groove textured gas film begins with a description of the film thickness with texture parameters and that of the textured sub-domains of the plan form. The finite-element method was thought to be the natural approach to treat the oblique geometry of inclined grooves; development of a finite-difference algorithm for non-orthogonal grids soon followed. In view of the multitude of design parameters, a thorough evaluation of design trade-offs via finite-element or finite-volume algorithms is at best very labor-intensive, since performance considerations invariably favor the largest possible number of grooves in the pattern. Narrow

groove analysis (NGA) makes available an efficient engineering tool to examine design options of textured gas bearings, especially since the threshold for stable operation cannot be taken for granted.

NGA recognizes two scales; texture-related issues are in the micro-scale of a representative angular width of the texture pattern. The macro-scale concerns the plan form description, subdivision of the bearing surface into interconnected textured and smooth annular sub-domains, and the bearing geometry, which is represented by a radial intersection of the bearing surface. The isothermal gas film approximation remains valid for both micro- and macro-scale analyses; albeit at different scales:

Scaling Factors

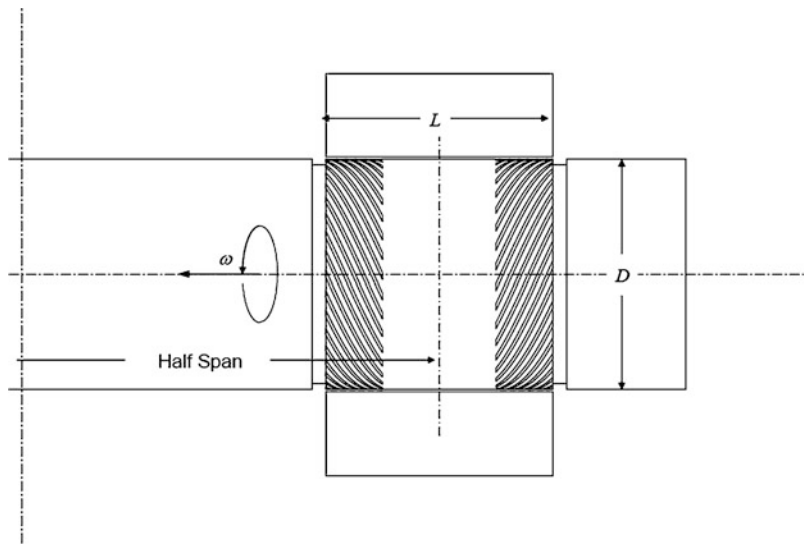
	Micro	Macro
Coordinates	$r\Delta\theta$	$R$
Film pressure	$\bar{p}(r)$	$p_a$
Flux vector	$\bar{p}rC$	$p_aRC$



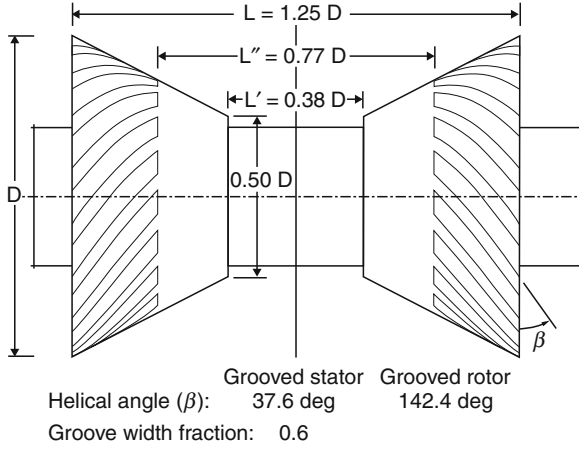
Gas Bearings with Narrow Inclined Grooves, Fig. 2 Spiral-groove thrust bearing plan forms

Micro-Scale Analysis

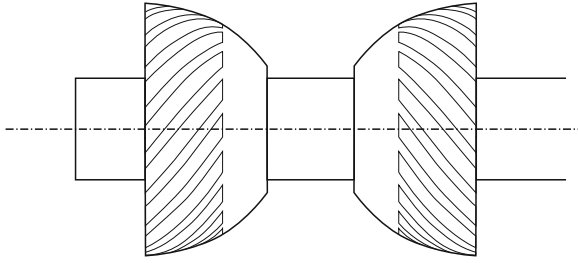
The micro-scale analysis deals with two alternative orthogonal Cartesian coordinates that are oblique to each other; see Fig. 1 for the description of corresponding base-vectors  $(\vec{i}, \vec{j})$  and  $(\vec{\lambda}, \vec{v})$ .  $\vec{i}$  is directed along the relative circumferential sliding motion of the smooth surface to the textured surface, while  $\vec{\lambda}$  is directed along the



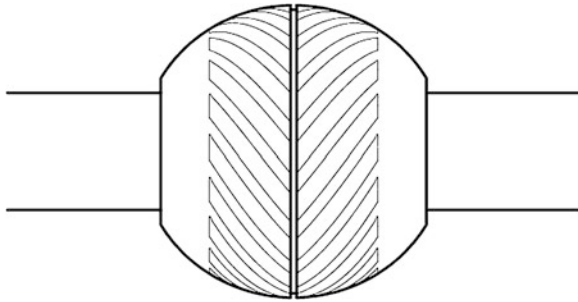
Gas Bearings with Narrow Inclined Grooves, Fig. 3 Journal bearing with helical grooves at shaft ends



Gas Bearings with Narrow Inclined Grooves, Fig. 4 Spiral grooved conical bearing



Gas Bearings with Narrow Inclined Grooves, Fig. 5 Spiral grooved hemisphere for omni-directional support



Gas Bearings with Narrow Inclined Grooves, Fig. 6 Free rotor spherical bearing

oblique step side of the texture. The locally scaled non-dimensional circumferential flux component is

$$\begin{aligned}\tilde{\Phi}_\theta|_{-\alpha < \theta < 0} &= \tilde{P}H_{\text{groove}} - \frac{\tilde{P}H_{\text{groove}}^3}{\tilde{\Lambda}} \frac{\partial \tilde{P}}{\partial \theta} \\ \tilde{\Phi}_\theta|_{0 < \theta < 1-\alpha} &= \tilde{P}H_{\text{land}} - \frac{\tilde{P}H_{\text{land}}^3}{\tilde{\Lambda}} \frac{\partial \tilde{P}}{\partial \theta}\end{aligned}\quad (1)$$

**Cross-Pattern Pressure Profiles.** The infinitesimal narrow groove analysis (INGA), as pioneered by Whipple, stipulates that the combination of any finite  $\Lambda$  with an infinitesimal  $\Delta\theta$  would yield a finite  $\partial\tilde{P}/\partial\tilde{\theta} = \tilde{\Lambda}^{-1}(\partial\tilde{P}/\partial\tilde{\theta})$ ; then (1) can be re-stated for the entire pattern span as

$$\tilde{\Phi} = \tilde{P}(\tilde{i}H - H^3\tilde{\nabla}\tilde{p}) \quad (2)$$

This is the INGA doctrine: texture-induced flow obeys the incompressible fluid film lubrication theory. For problems with rotational symmetry,  $\tilde{p}$  is periodic and circumferential integration of (2) yields a triangular profile of  $\tilde{p}$  with discontinuity in both  $\partial\tilde{P}/\partial\tilde{\theta}$  and  $\tilde{\Phi}_\theta$ , as summarized below:

$$\begin{aligned}\delta\tilde{\theta}_{\text{groove}} & \quad \delta\tilde{\theta}_{\text{land}} & \quad \text{sgn}\left\{\frac{\partial\tilde{p}}{\partial\tilde{\theta}}\right\} & \quad \tilde{\Phi}_\theta \\ \alpha & \quad 1-\alpha & \quad -\text{sgn}\left\{\tilde{\theta}\right\} & \quad H - H^3\frac{\partial\tilde{p}}{\partial\tilde{\theta}}\end{aligned}\quad (3)$$

Whipple treated the thrust bearing problem with rotational symmetry; here, to allow for radial eccentricity, cross-pattern periodicity of  $\tilde{p}$  is supplemented by  $(\Lambda\tilde{r})^{-1}\Delta\theta(\partial\tilde{P}/\partial\tilde{\theta})$ , yielding

$$\partial\tilde{P}/\partial\tilde{\theta} = \partial\tilde{p}/\partial\tilde{\theta} + (\Lambda\tilde{r}^2)^{-1}(\partial\tilde{P}/\partial\tilde{\theta}) \quad (4)$$

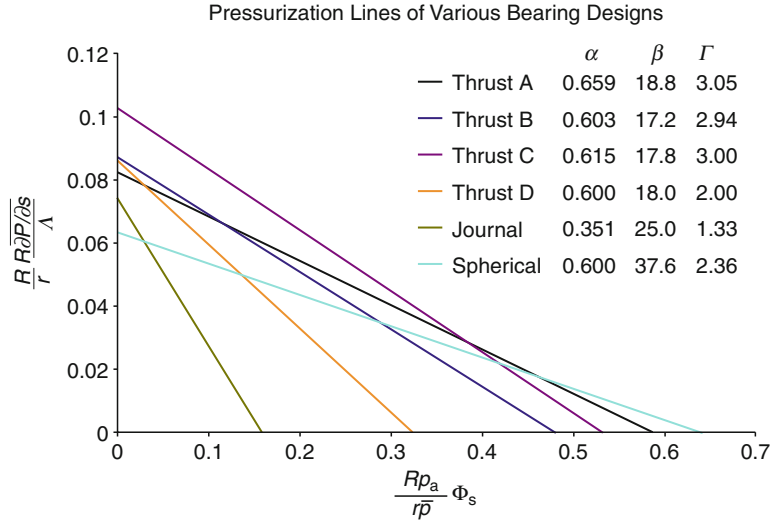
$\tilde{p}$  remains periodic across the texture pattern.

**Cross-Step Continuity.** INGA also results in the invariance of the transverse flux in each sub-interval of the texture pattern; they are determined from cross-step continuity conditions of the longitudinal pressure gradient  $\tilde{\lambda} \cdot \tilde{\nabla}\tilde{p}$  and of the normal film flux  $\tilde{v} \cdot \tilde{\Phi}$  simultaneously. They are

$$\begin{aligned}\tilde{\lambda} \cdot \tilde{\nabla}\tilde{p} &= (\partial\tilde{p}/\partial\tilde{s}) \sin\beta + \left(\partial\tilde{p}/\partial\tilde{\theta}\right) \cos\beta \\ \tilde{v} \cdot \tilde{\Phi} &= -H^3(\partial\tilde{p}/\partial\tilde{s}) \cos\beta \\ &\quad - \left[H - H^3\left(\partial\tilde{p}/\partial\tilde{\theta}\right)\right] \sin\beta\end{aligned}\quad (5)$$

Implementation of these cross-step continuity requirements repeatedly encounters the following algebraic operations:

$$\begin{aligned}\Delta(\ ) &\equiv (\ )_{\text{groove}} - (\ )_{\text{land}} \\ \overline{(\ )} &\equiv \alpha(\ )_{\text{groove}} + (1-\alpha)(\ )_{\text{land}} \\ \overline{\overline{H^3}} &\equiv H_{\text{groove}}^3 H_{\text{land}}^3\end{aligned}\quad (6)$$



Gas Bearings with Narrow Inclined Grooves, Fig. 7 Typical pressurization lines

A single formula is found for the invariant  $\partial \tilde{p}/\partial \tilde{s}$  in either the groove or the land with the respective  $H^{-3}$  in the coefficient of  $\text{ctn}^2 \beta$ :

$$\frac{\partial \tilde{p}}{\partial \tilde{s}} = \left\{ \left( \overline{H^{-3}} + H^{-3} \text{ctn}^2 \beta \right) \frac{\overline{H^3}}{\Delta H^3} \frac{\delta \tilde{p}}{\alpha(1-\alpha)} - \frac{\Delta H}{\Delta H^3} + \frac{\partial \tilde{p}/\partial \tilde{\theta}}{\Lambda \bar{r}^2} \right\} \tan \beta \quad (7)$$

The invariant  $\tilde{\Phi}_s$  is found by multiplying the corresponding  $-H^3$  into (7).

Cross-pattern averaging of  $\partial \tilde{p}/\partial \tilde{s}$  and  $\tilde{\Phi}_s$  renders macro-scale transverse pressure gradient and flux in local scales:

$$\begin{aligned} \overline{\partial \tilde{p}/\partial \tilde{s}} &\equiv \int_{-\alpha}^{1-\alpha} (\partial \tilde{p}/\partial \tilde{s}) d\tilde{\theta} \\ &= -\tan \beta \left( \frac{\Delta H}{\Delta H^3} - \frac{\partial \tilde{p}/\partial \tilde{\theta}}{\Lambda \bar{r}^2} \right) \\ &\quad + \frac{\overline{H^3} \overline{H^{-3}}}{\Delta H^3 \alpha(1-\alpha) \sin \beta \cos \beta} \delta \tilde{p} \end{aligned} \quad (8)$$

$$\begin{aligned} \overline{\tilde{\Phi}_s} &\equiv \int_{-\alpha}^{1-\alpha} \tilde{\Phi}_s d\tilde{\theta} = \overline{\tilde{\Phi}_s} / (\bar{r} \bar{P}) \\ &= \overline{H^3} \tan \beta \left( \frac{\Delta H}{\Delta H^3} - \frac{\partial \tilde{p}/\partial \tilde{\theta}}{\Lambda \bar{r}^2} \right) \\ &\quad - \frac{\overline{H^3} [(\overline{H^3} \overline{H^{-3}} - 1) \sin^2 \beta + 1]}{\Delta H^3 \alpha(1-\alpha) \sin \beta \cos \beta} \delta \tilde{p} \end{aligned} \quad (9)$$

Equations (8) and (9) are linear parametric equations of  $\overline{\partial \tilde{p}/\partial \tilde{s}}$  and  $\overline{\tilde{\Phi}_s}$  in terms of  $\delta \tilde{p}$  and  $\partial \tilde{p}/\partial \tilde{\theta}$ . In the absence of radial eccentricity,  $\overline{\partial \tilde{p}/\partial \tilde{s}}$  can be graphed against  $\overline{\tilde{\Phi}_s}$  with  $\delta \tilde{p}$  varied between the intercept values

$$\begin{aligned} \delta \tilde{p}|_{\overline{\tilde{\Phi}_s}=0} &= \frac{\beta \overline{H^3} \Delta H \alpha (1-\alpha)}{\overline{H^3} [(\overline{H^3} \overline{H^{-3}} - 1) \sin^2 \beta + 1]} \\ \delta \tilde{p}|_{\partial \tilde{p}/\partial \tilde{s}=0} &= \frac{\sin^2 \Delta H \alpha (1-\alpha)}{\overline{H^3} \overline{H^{-3}}} \end{aligned} \quad (10)$$

This is a downward-inclined straight line connecting the intercepts that are functions of the texture parameters  $(\alpha, \beta, \Gamma)$ . Figure 7 shows several typical pressurization lines of various bearings. Since the two intercept points are synthesized from three texture parameters, a textured bearing is usually designed for some chosen objective with the possibility of optimized performance.

Elimination of  $\delta \tilde{p}$  between (8) and (9) yields the macro-scale transverse flux with allowance for a non-vanishing  $\partial P/\partial \theta$ :

$$\begin{aligned} \left( \bar{r} \bar{P} \right) \overline{\tilde{\Phi}_s} &= -(\overline{H^3} - 1/\overline{H^{-3}}) \sin \beta \cos \beta (\Lambda \bar{r}^2)^{-1} \partial \tilde{p}/\partial \tilde{\theta} \\ &\quad - (\overline{H^3} \sin^2 \beta + \cos^2 \beta / \overline{H^{-3}}) (\Lambda \bar{r})^{-1} \partial \tilde{p}/\partial \tilde{s} \\ &\quad + (\overline{H^3} - 1/\overline{H^{-3}}) \sin \beta \cos \beta (\Delta H / \Delta H^3) \end{aligned} \quad (11)$$

### Macro-Scale Analysis

Cross-pattern averaged  $(\overline{\tilde{\Phi}}, \overline{\tilde{\nabla} \tilde{p}})$ , as defined by (4), (8), and (9) are non-dimensional variables in micro-scaling.



They are related to their respective non-dimensional macro-scale counterparts as follows:

$$\begin{aligned}\bar{\Phi}(\theta, \bar{r}) &= \bar{i} \bar{\Phi}_\theta + \bar{j} \bar{\Phi}_s = \bar{r} \bar{P} \left( \bar{i} \bar{\Phi}_\theta + \bar{j} \bar{\Phi}_s \right) \\ \bar{P}(\theta, \bar{r}) &= (\bar{p}/p_a) \bar{P}\end{aligned}\quad (12)$$

Thus written, they are combined to form the two-dimensional fluid film continuity condition:

$$\bar{\nabla} \cdot \bar{\Phi} + \frac{\partial}{\partial \tau} (\bar{P} \bar{H}) = 0 \quad (13)$$

Bearing geometry is defined by the function that describes its radial plane cross-section  $dr/ds$  given in (14) and the corresponding two-dimensional formulas for the divergence of film flux are listed in (15):

Bearing	$dr/ds$	
Journal	0	
Thrust	1	
Conical	$\sin \frac{1}{2} \gamma$	(14)
Spherical	$\cos \varphi$	

Bearing	$\bar{\nabla} \cdot \bar{\Phi}$	
Journal	$\partial \Phi_\theta / \partial \theta + \partial \Phi_s / \partial \bar{s}$	
Thrust	$\partial \Phi_\theta / \partial \theta + \frac{\partial(\bar{r} \Phi_s)}{\bar{r} \partial \bar{r}}$	(15)
Conical	$\partial \Phi_\theta / \partial \theta + \frac{\partial(\bar{r} \Phi_s)}{\bar{r} \partial \bar{s}}$	
Spherical	$\frac{\partial \Phi_\theta}{\partial \theta} + \sin \varphi \frac{\partial(\bar{r} \Phi_s)}{\bar{r} d\bar{r}}$	

The formulas of  $\bar{\Phi}$  and  $\bar{P}$  are restated below with  $\delta \bar{p}$  and  $\partial \bar{p} / \partial \bar{s}$  of each pattern subdivision expressed in macro-scale convention. They are

$$\begin{aligned}& \frac{\bar{H}^3 \bar{H}^{-3} \delta \bar{p}}{\Delta H^3 \alpha (1 - \alpha) \sin \beta \cos \beta} \\ &= (\Lambda \bar{r})^{-1} \bar{P} \left[ (\partial \bar{P} / \partial \bar{s}) - \tan \beta \bar{r}^{-1} (\partial \bar{P} / \partial \theta) \right] \\ &+ \tan \beta (\Delta H / \Delta H^3)\end{aligned}\quad (16)$$

$$\begin{aligned}\frac{\partial \bar{p}}{\partial \bar{s}} &= (\sin^2 \beta + \cos^2 \beta H^{-3} / \bar{H}^{-3}) \frac{\bar{P}(\partial \bar{P} / \partial \bar{s})}{\Lambda \bar{r}} \\ &+ \frac{(H^{-3} - \bar{H}^{-3}) \Delta H \sin \beta \cos \beta}{\bar{H}^{-3} \Delta H^3} \\ &+ (H^{-3} / \bar{H}^{-3} - 1) \sin \beta \cos \beta \frac{\bar{P}(\partial \bar{P} / \partial \theta)}{\Lambda \bar{r}^2}\end{aligned}\quad (17)$$

A macro-scale circumferential flux law is derived from INGA parameters by performing transverse cross-pattern

integration of  $\tilde{\Phi}_\theta$  with allowance for the profile of  $\partial \tilde{p} / \partial \tilde{\theta}$  in each sub-division. Each of these flux laws is a three-term formula; they, respectively, represent the effects of texture-wedge, longitudinal pressure gradient, and orthogonal pressure gradient with coefficients that are functions of INGA parameters:

$$\begin{aligned}\bar{\Phi}_\theta &\equiv \bar{r} \bar{P} \int_{-(1-\alpha)}^{\alpha} \tilde{\Phi}_\theta d(\tilde{s} / \tan \beta) \\ &= \bar{r} \bar{P} (\bar{H} + c_{\theta, \Gamma}) \\ &- (\bar{H}^3 + c_{\theta, \Phi}) \frac{\bar{P}(\partial \bar{P} / \partial \theta)}{\Lambda \bar{r}} + c_{\theta, X} \frac{\bar{P}(\partial \bar{P} / \partial \bar{s})}{\Lambda} \quad (18) \\ c_{\theta, \Gamma} &= -\alpha (1 - \alpha) \sin^2 \beta \Delta H \\ c_{\theta, \Phi} &= -\alpha (1 - \alpha) \sin^2 \beta \Delta H^3 \\ c_{\theta, X} &= \Delta H^3 \alpha (1 - \alpha) \sin \beta \cos \beta\end{aligned}$$

$$\begin{aligned}\bar{\Phi}_s &\equiv \bar{j} \cdot \bar{\Phi}(\theta, \bar{r}) = \bar{r} \bar{P} \bar{\Phi}_s \\ &= (\bar{r} \bar{P}) c_{s, \Gamma} - c_{s, X} \frac{\bar{P} \partial \bar{P} / \partial \theta}{\Lambda \bar{r}} - c_{s, \Phi} \frac{\bar{P} \partial \bar{P} / \partial \bar{s}}{\Lambda} \quad (19) \\ c_{s, \Gamma} &= (\bar{H}^3 - 1 / \bar{H}^{-3}) \sin \beta \cos \beta (\Delta H / \Delta H^3) \\ c_{s, X} &= (\bar{H}^3 - 1 / \bar{H}^{-3}) \sin \beta \cos \beta \\ c_{s, \Phi} &= \bar{H}^3 \sin^2 \beta + \cos^2 \beta / \bar{H}^{-3}\end{aligned}$$

The above formulas reduce to the standard isothermal gas film flux law for  $\Gamma \rightarrow 0$ . A generic plan form of two sub-domain annuli separated by an internal boundary at  $\bar{s}_m$  is featured in many designs.

Because textured bearings are inherently stiff, their macro-scale analysis can be conveniently treated as the sum of a steady-state solution, which is one-dimensional due to rotational symmetry, and a perturbation solution; the latter is separated into three degrees of freedom (DOF) with respect to axial, radial, and angular displacements of the supported rotor. Accordingly, the non-dimensional film thickness can be expressed as Jacobian perturbations

$$\begin{aligned}\frac{\partial H}{\partial(\varepsilon_z, \varepsilon_r, \varepsilon_\theta)} &= (H_z, H_r, H_\theta) \\ \delta H &= H_z \varepsilon_z + H_r \varepsilon_r + H_\theta \varepsilon_\theta\end{aligned}\quad (20)$$

Subscript “0” will designate the steady-state macro-scale solution that correspond to nil  $(\varepsilon_z, \varepsilon_r, \varepsilon_\theta)$ .

Steady-State Solution. The macro-scale steady-state problem corresponds to the one-dimensional case without the time dependent term; (13) is reduced to

$$\frac{\partial}{\partial \bar{s}} (\bar{r} \bar{\Phi}_s) = 0 \quad (21)$$

$H_{\text{land}} = 1$  is assigned for steady-state operation.  $P_0$  is used as the simplified symbol for the macro-scale film pressure. The boundary conditions are

$$P_0(\bar{s}_{\text{entrance}}) = P_0(\bar{s}_{\text{exit}}) = 1 \quad (22)$$

An unspecified (if not nil)  $\Phi_s(\bar{s}_m)$  continues across the internal boundary between the sub-domains of the bearing plan form.

A general solution of (21) for the span bracketed between  $(\bar{s}_{\text{entrance}}, \bar{s}_m)$  is the through-flow invariant formula

$$\bar{r}\bar{\Phi}_s(\bar{s}) = \text{Invariant} = \Psi \quad (23)$$

Setting  $\partial\bar{P}/\partial\theta = 0$ , (19) becomes

$$\Psi = \bar{r}^2 P_0 (c_{s,\Gamma} - c_{s,\Phi} \frac{dP_0}{\Lambda \bar{r} d\bar{s}}) \quad (24)$$

For a bearing with non-vanishing through flow, (24) would be calculated from the inlet ambient condition, with an assumed flow invariant to reach  $\bar{s}_m$  to establish  $P_{0,m}$ ; the second sub-domain would be a smooth annulus with  $c_{s,\Gamma} = 0$  and  $c_{s,\Phi} = 1$ , so that (24) is reduced to

$$\Psi = -\frac{P_0 dP_0}{\Lambda d\bar{s}/\bar{r}} \quad (25)$$

Integrating, it is found

$$\frac{P_{0,m}^2 - P_0^2}{2\Lambda\Psi} = \int_{\bar{s}_m}^{\bar{s}} \frac{d\bar{s}}{\bar{r}} \quad (26)$$

The correct value of the flow invariant would allow (26) to satisfy the exit ambient condition.

**Herringbone Bearings.** All bearing types can utilize a herringbone plan form to suppress the through-flow such that texture inclination reverses at  $\bar{s}_m$ . Equation (24) is reduced to

$$\left. \frac{dP_0}{\Lambda \bar{r} d\bar{s}} \right|_{\text{herringbone}} = -\text{sgn}\{\bar{s} - \bar{s}_m\} (|c_{s,\Gamma}|/c_{s,\Phi}) \quad (27)$$

The profile on either side of the unknown  $\bar{s}_m$  is

$$(\Lambda |c_{s,\Gamma}|/c_{s,\Phi})^{-1} (P_0 - 1) = \int_{\bar{s}}^{\bar{s}_m} \bar{r} d\bar{s} \quad (28)$$

Applying the ambient conditions  $P_0|_{\bar{s}_{\text{ambient}}} = 1$ , it is found

$$\begin{aligned} P_{0,m} &= 1 - \Lambda (|c_{s,\Gamma}|/c_{s,\Phi}) \int_{\bar{s}_1}^{\bar{s}_m} \bar{r} d\bar{s} \\ &= 1 + \Lambda (|c_{s,\Gamma}|/c_{s,\Phi}) \int_{\bar{s}_1}^{\bar{s}_m} \bar{r} d\bar{s} \end{aligned} \quad (29)$$

Making use of (14) for various bearing types, one obtains paired  $(\bar{s}_m, P_{0,m})$  of the herringbone plan forms as listed below:

Bearing	$\int \bar{r} d\bar{s}$	$\bar{s}_m$
Journal	$\bar{r}\bar{z}$	0
Thrust	$\frac{1}{2}\bar{r}^2$	$\sqrt{\frac{1}{2}(1 + \bar{r}_i^2)}$
Conical	$\frac{1}{2}\bar{r}^2$	$\sqrt{\frac{1}{2}(1 + \bar{r}_i^2)}$
Spherical	$\frac{1}{\sin \frac{1}{2}\gamma}$	$\frac{\sin \frac{1}{2}\gamma}{\cos^{-1} \frac{\cos \varphi_i + \cos \varphi_1}{2}}$

$$\quad (30)$$

Bearing	$\frac{P_{0,m}-1}{\Lambda (c_{s,\Gamma}/c_{s,\Phi})}$
Journal	$L/D$
Thrust	$\frac{1}{4}(1 - \bar{r}_i^2)$
Conical	$\frac{1}{4}(1 - \bar{r}_i^2)$
Spherical	$\frac{1}{2}(\cos \varphi_i - \cos \varphi_1)$

$$\quad (31)$$

**Transverse Through-Flow.** Equation (24) can still be solved in closed form for the journal bearing even if the through-flow does not vanish; one finds

$$\begin{aligned} (P_{0,m} - 1) + (\Psi/c_{s,\Gamma}) \ln \left( \frac{P_{0,m} - \Psi/c_{s,\Gamma}}{1 - \Psi/c_{s,\Gamma}} \right) \\ = \Lambda (c_{s,\Gamma}/c_{s,\Phi}) (L/D + \bar{z}_m) \end{aligned} \quad (32)$$

In the smooth sub-domain  $\bar{z} > \bar{z}_m$ , (26) can be used with the right-hand side equated to the axial length of the exit end measured from the internal boundary:

$$P_{0,m}^2 - 1 = 2\Lambda\Psi (L/D - \bar{z}_m) \quad (33)$$

For any specified  $-L/D < \bar{z}_m < L/D$ , (32) and (33) can be conjunctively satisfied through numerical iterations.

Equation (24) is mathematically equivalent to the air-lubricated taper-slider problem treated by Harrison. It has an asymptotic limiting solution:

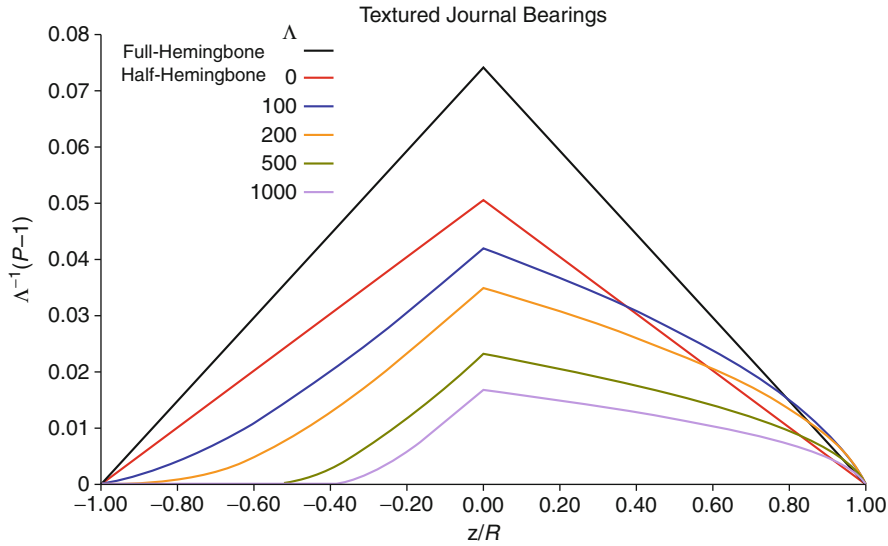
$$\lim_{\Lambda \rightarrow \infty} \left\{ \Psi \rightarrow c_{s,\Gamma} \quad P_0|_{\bar{z} < \bar{z}_m} = 1 \right\} \quad (34)$$

$P_0$  steeply rises to meet  $P_{0,m}$  in a narrow region of  $O\{\Lambda^{-1}\}$ , satisfying a simple extrapolation formula:

$$\lim_{\Lambda \rightarrow \infty} \Lambda^{-1} (P_{0,m} - 1) = \sqrt{2\Lambda^{-1} c_{s,\Gamma} (L/D - \bar{z}_m)} \quad (35)$$

Setting  $\bar{z}_m = 0$  to terminate surface texturing, one obtains a half-herringbone journal bearing. Its pressure profiles at various bearing numbers are compared to that of the full-herringbone bearing in Fig. 8, using same texture parameters previously listed in Fig. 7.

With through-flow fully suppressed, the full-herringbone operates at the left threshold of the



Gas Bearings with Narrow Inclined Grooves, Fig. 8 Pressure profiles of journal bearing

pressurization line; the pressure profile is triangular and there is no  $\Lambda$ –dependence. The half-herringbone journal bearing shows considerable leakage, as evidence by about 32% reduction in the peak pressure in the incompressible profile ( $\Lambda \rightarrow 0$ ), which also has a triangular shape. As  $\Lambda$  increases, the peak is moderately reduced and the profile becomes concave in the textured left side, accompanied by convexity in the smooth right side. At  $\Lambda = 100$ , transverse through-flow is over 82% of the right threshold of the pressurization line; however, peak pressure remains at a respectable magnitude. At  $\Lambda = 200$ , transverse through-flow is over 99% of the right threshold, while the pressure peak is still considerable. So far as the overall area under the pressure profile is concerned, there is not much change up to  $\Lambda = 200$ . Flattening of the pressure profile near the left entrance becomes prominent. As  $\Lambda$  further increases, the asymptotic condition indicated by (34) clearly sets in.

There is no closed form solution of (24) for other bearing types with a non-vanishing  $\Psi$ . In view of the existence of a large  $\Lambda$  asymptotic behavior according to (34), one desires a robust numerical treatment that emulates Harrison's formula while retaining  $\bar{r}$ -dependent coefficients in the right-hand side of (24); with the aid of a mean value approximation in the computation interval from  $\bar{s}_{k-1}$  to  $\bar{s}_k$  for  $P_0 \approx P_0|_{\text{mean}} \equiv P_0|_{\bar{s}_{k-1}} + \frac{1}{2}\Delta P_0$ , an exact integral from  $\bar{s}_{k-1}$  to  $\bar{s}_k$  can be derived as

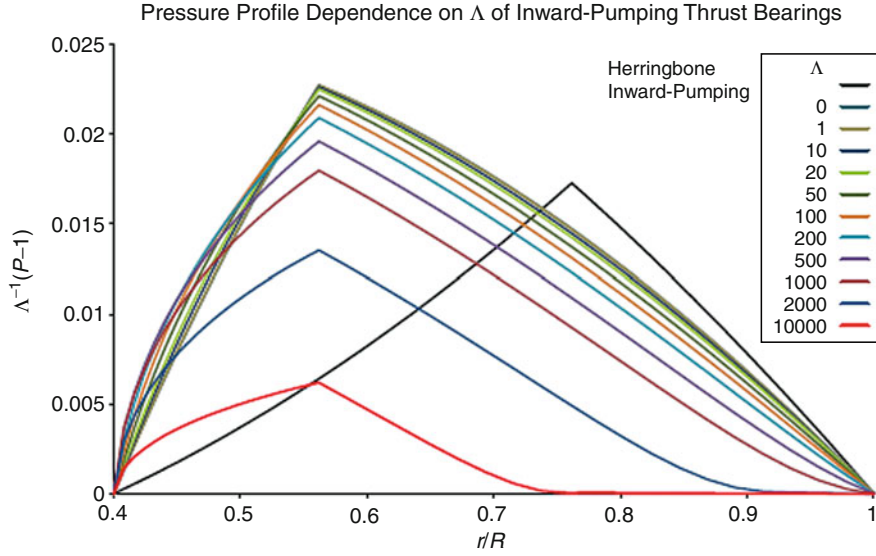
$$\Lambda^{-1}\Delta P_0 = \frac{c_{s,\Phi}}{c_{s,\Gamma}} \int_{\bar{s}_{k-1}}^{\bar{s}_k} \bar{r} d\bar{s} - \frac{\Psi/c_{s,\Gamma}}{P_0|_{\text{mean}}} \int_{\bar{s}_{k-1}}^{\bar{s}_k} \frac{d\bar{s}}{\bar{r}} \quad (36)$$

The integral of  $1/\bar{r}$  with respect to  $\bar{s}$  brings in the geometrical influence on the transverse through-flow of each bearing type:

Bearing	$\int d\bar{s}/\bar{r}$	(37)
Thrust	$\ln \bar{r}$	
Conical	$\csc \frac{1}{2}\gamma \ln \bar{r}$	
Spherical	$\frac{1}{2} \ln \frac{1-\cos \varphi}{1+\cos \varphi}$	

Equation (36) yields a second order polynomial of  $\Delta P_0$  that can be calculated as a quadratic root;  $P_0|_{\bar{s}_k}$  is then calculated by adding  $\Delta P_0$  to  $P_0|_{\bar{s}_{k-1}}$ .

The competing roles of pressurization and flow restriction caused by continuous reduction of  $\bar{r}$  are clearly revealed in the pressure profiles of a herringbone thrust bearing and an inward-pumping thrust bearing, illustrated in Fig. 9. All profiles are convex both in the outer annulus and in the inner annulus. The peak pressure of the herringbone thrust bearing is considerably lower than those of the inward-pumping design for  $\Lambda \leq 1000$  the herringbone thrust bearing is handicapped by the relatively large  $\bar{r}_m = 0.76$ , which is required to balance inward- and outward-pumping herringbone annuli. The much smaller  $\bar{r}_m = 0.56$  of the inward-pumping design allows pressurization to reach a higher peak despite the presence of through flow, which is quite effectively restricted by the smaller radii of the inner annulus. The pressure profiles hardly vary up to  $\Lambda = 20$ . Overall area under the profile compares favorably to that of the full-herringbone plan form up to  $\Lambda = 1000$ . A hint of the asymptotic condition is seen at  $\Lambda = 2000$  and it is in full



Gas Bearings with Narrow Inclined Grooves, Fig. 9 Pressure profiles of inward-pumping thrust bearing

force at  $\Lambda = 10000$ . It thus can be surmised that for bearing types with a varying cylindrical radius ( $dr/ds \neq 0$ ) the inward-pumping plan form is generally the preferred design.

Figures 8 and 9 illustrate that inclined-groove texturing is inherently suited to large  $\Lambda$  operation.

**Perturbation analysis.** Perturbation analysis is concerned with departure from steady-state operation of the bearing. Consider

$$(\bar{\Phi}, \bar{P}) = (\bar{\Phi}_0, P_0) + \delta(\bar{\Phi}, P) \quad (38)$$

The steady-state result is defined by the existence of a flow invariant  $\Psi = \bar{r}\bar{\Phi}_{s,0}$  and the associated  $P_0(\bar{s})$  that satisfies (24). Formal substitution of (38) into (13) yields the condition of fluid film continuity for a perturbed state:

$$\frac{\partial(\bar{r}\delta\bar{\Phi}_s)}{\bar{r}\partial\bar{s}} + \frac{\partial\delta\bar{\Phi}_\theta}{\bar{r}\partial\theta} + \frac{\partial(P_0\delta\bar{H} + \bar{H}\delta P)}{\frac{1}{2}\partial\tau} = 0 \quad (39)$$

Jacobian derivatives are used to account for  $H$ -dependence of the flux coefficients. They are

$$\begin{aligned} & (c_{\theta,X;H}, c_{s,\Gamma;H}, c_{s,\Phi;H}) \\ &= \frac{\partial(c_{\theta,X}, c_{s,\Gamma}, c_{s,\Phi})}{\partial(\Delta H^3, \bar{H}^3, \bar{H}^{-3})} \frac{\partial(\Delta H^3, \bar{H}^3, \bar{H}^{-3})}{H} \quad (40) \\ & \partial\Delta H^3/\partial H = 3\Delta H^2, \quad \partial\bar{H}^{\pm 3}/\partial H = \pm 3\bar{H}^{\pm 3-1} \end{aligned}$$

The first term on the left-hand side of (39) is

$$\begin{aligned} \frac{\partial(\bar{r}\delta\bar{\Phi}_s)}{\bar{r}\partial\bar{s}} &= \frac{\partial}{\bar{r}\partial\bar{s}} \left\{ \bar{r}^2 \left[ c_{s,\Gamma;H} - \frac{c_{s,\Phi;H}}{c_{s,\Phi}} \left( c_{s,\Gamma} - \frac{\Psi}{P_0\bar{r}} \right) \right] P_0\delta H \right. \\ & \quad \left. - \frac{P_0}{\Lambda} \left( c_{s,X} \frac{\partial\delta P}{\partial\theta} + c_{s,\Phi} \frac{\bar{r}\partial\delta P}{\partial\bar{s}} \right) + \frac{\Psi\delta P}{P_0} \right\} \quad (41) \end{aligned}$$

**Axial Perturbation.** The bearing film thickness retains rotational symmetry under axial perturbation. Film thickness perturbation is  $\delta H = H_z e_z$ ,  $\partial/\partial\theta$  is suppressed for both  $P_0$  and  $\bar{\Phi}_0$ . A stability study of the rotor bearing system (► [Self-Excited Gas Bearing Instabilities](#)) requires full-frequency compilation of  $\Omega$ -dependent bearing reaction to rotor displacement.

The governing equation for the most general axial perturbation problem is found as

$$\begin{aligned} & j\Omega(\bar{H}\delta P + P_0 H_z e_z) = \\ & \frac{d}{\bar{r}d\bar{s}} \left[ c_{s,\Phi} \frac{P_0}{\Lambda} \frac{\bar{r}d\delta P}{d\bar{s}} - \frac{\Psi}{P_0} \delta P - \frac{c_{s,\Phi;H}}{c_{s,\Phi}} \bar{r}\Psi H_z e_z \right. \\ & \quad \left. - P_0 \bar{r}^2 \left( c_{s,\Gamma;H} - c_{s,\Gamma} \frac{c_{s,\Phi;H}}{c_{s,\Phi}} \right) H_z e_z \right] \quad (42) \end{aligned}$$

Film thickness perturbation function is

Bearing	$H_z$	
Thrust	$-1$	
Conical	$-\sin \frac{1}{2}\gamma$	(43)
Spherical	$-\cos \varphi = -\sqrt{1 - \bar{s}^2}$	



Spatial differentiation is reverted to the ordinary  $\bar{s}$ - derivative since there is no circumferential variation and multiplication by  $j\Omega$  replaces  $\tau$ - derivative. The null boundary condition  $\delta P = 0$  applies at ambient edges. Equation (42) can be readily computed to yield the complex perturbation pressure

$$\delta P_z = (u_z + jv_z)\varepsilon_z \quad (44)$$

Integration of the complex perturbed film pressure over the bearing area yields the complex axial reaction impedance; in the non-dimensional form, this is

$$\begin{aligned} \frac{\delta F_z}{\varepsilon_z \pi R^2 p_a} &= - \int_{\text{plan form}} \delta P(\bar{s}) H_z \bar{r} d\bar{r} \\ &= (U_z + jV_z) \end{aligned} \quad (45)$$

For a stability study, a complete set of  $U_z(\Omega)$ ,  $V_z(\Omega)$  would be compiled. Because all coefficients in the right-hand side of (42) are real,  $U_z + jV_z$  is a conjugate function of  $\Omega$  that has the asymptotic property

$$\lim_{|\Omega| \rightarrow \infty} V_z = o\{\Omega^{-1}\} \quad (46)$$

*Radial Perturbation.* Film thickness perturbation caused by a radial rotor displacement is

$$\delta H = H_r \varepsilon_r \cos \theta = \varepsilon \cos \theta \quad (47)$$

$\varepsilon$  is an unspecified infinitesimal eccentricity ratio.  $\partial \delta P / \partial \theta$  in  $\bar{\Phi}_s$  according to (19) is to be retained for substitution into (39). Simple harmonic dependence on  $\theta$  is expected of  $\delta P$  in view of (47). Because  $\partial(\delta P \bar{H} + P_0 \delta H) / \partial \theta$  in  $\partial \delta \bar{\Phi}_\theta / \partial \theta$  is paired with  $\partial(\delta P \bar{H} + P_0 \delta H) / (\frac{1}{2} \partial \tau)$  in (39), it is advantageous to introduce the moving coordinate

$$\bar{\theta}_\pm \equiv \theta \mp 2|\Omega|\tau \quad (48)$$

Thus

$$\begin{aligned} \partial / \partial \theta &= \left( \partial / \partial \bar{\theta}_\pm \right) \left( \partial \bar{\theta}_\pm / \partial \theta \right) = \left( \partial / \partial \bar{\theta}_\pm \right) \\ \partial / \partial \tau &= \left( \partial / \partial \bar{\theta}_\pm \right) \left( \partial \bar{\theta}_\pm / \partial \tau \right) = \mp 2|\Omega| \left( \partial / \partial \bar{\theta}_\pm \right) \end{aligned} \quad (49)$$

Consequently

$$\frac{\partial}{\partial \theta} + \frac{\partial}{\partial \tau} = (1 \mp 2|\Omega|) \frac{\partial}{\partial \bar{\theta}_\pm} \quad (50)$$

$\Omega = \pm |\Omega|$  are independent conditions that can be related to two independent linear displacement DOFs of

the rotor to its rotation axis. Equation (39) can now be converted into the rotating coordinate system; all  $\partial / \partial \theta$  operations in  $\bar{\nabla} \cdot \delta \bar{\Phi}$  are replaced by  $\partial / \partial \bar{\theta}$  and whirl rate weighted  $\bar{\theta}$ - derivative is substituted for explicit  $\tau$ - derivative by virtue of (49). Thus the steady-whirl problem is governed by

$$\frac{\partial(\bar{r} \delta \bar{\Phi}_s)}{\bar{r} \partial \bar{s}} + \frac{\partial \delta \bar{\Phi}_\theta}{\bar{r} \partial \bar{\theta}} - 2\Omega \frac{\partial(P_0 \delta \bar{H} + \bar{H} \delta P)}{\partial \bar{\theta}} = 0 \quad (51)$$

The convention of implied complex exponential factor  $\exp\{j\bar{\theta}\}$  can be used with  $(\delta H, \delta P)$  to describe their simple-harmonic dependence on  $\bar{\theta}$ . The whirl perturbation equation is

$$\begin{aligned} & - \frac{\partial}{\bar{r} \partial \bar{s}} \left( \frac{P_0}{\Lambda} \bar{r} c_{s,\Phi} \frac{\partial \delta P}{\partial \bar{s}} + \frac{\Psi \delta P}{P_0} \right) \\ & + j \left[ \frac{\partial}{\bar{r} \partial \bar{s}} \left( c_{s,X} \frac{P_0}{\Lambda} \delta P \right) + c_{\theta,\Gamma} + \frac{c_{s,\Gamma} c_{\theta,X}}{c_{s,\Phi}} \right] \delta P \\ & + (\bar{H}^3 + c_{\theta,\Phi}) \frac{P_0}{\Lambda \bar{r}^2} \delta P + j(1 - 2\Omega) \bar{H} \delta P \\ & + j \left( c_{\theta,X} \frac{P_0}{\Lambda} \frac{\partial \delta \bar{P}}{\partial \bar{s}} - \frac{c_{\theta,X}}{c_{s,\Phi}} \frac{\Psi}{P_0 \bar{r}^2} \right) \delta P \\ & + \frac{\partial}{\bar{r} \partial \bar{s}} \left\{ \left[ \bar{r}^2 \left( c_{s,\Gamma,H} - \frac{c_{s,\Gamma} c_{s,\Phi,H}}{c_{s,\Phi}} \right) \right. \right. \\ & \quad \left. \left. + \frac{c_{s,\Phi,H}}{c_{s,\Phi}} \frac{\Psi}{P_0 \bar{r}} \right] P_0 \varepsilon \right\} \\ & + j \left( \frac{c_{s,\Gamma} c_{\theta,X,H}}{c_{s,\Phi}} - \frac{c_{\theta,X,H}}{c_{s,\Phi}} \frac{\Psi}{P_0 \bar{r}^2} \right) P_0 \varepsilon \\ & + j(1 - 2\Omega) P_0 \varepsilon = 0 \end{aligned} \quad (52)$$

Again, spatial differentiation has reverted to the ordinary  $\bar{s}$ - derivative and the implied complex exponential coefficient represents a simple harmonic dependence on  $\bar{\theta}$  with phase indication. Numerical quadrature of (52) with respect to  $\bar{s}$ , with null boundary conditions  $\delta P$  at ambient edges, yields the complex perturbation pressure in response to rotor displacement  $(\varepsilon_r, \varepsilon_\theta)$  with the appropriate dependence on  $\bar{s}$ :

$$\delta P(\bar{s}) = (u + jv)\varepsilon \quad (53)$$

Perturbation solutions with radial eccentricity also obey the asymptotic condition

$$\lim_{\text{fixed } \Lambda, |\Omega| \rightarrow \infty} v = o\{\Omega^{-1}\} \quad (54)$$

In the case of cylindrical whirl, of interest is bearing reaction to a rotor displacement of amplitude  $\varepsilon_r$ .

The corresponding film thickness perturbation is, respectively,

$$H_r \quad \begin{array}{ccc} \text{Journal} & \text{Conical} & \text{Spherical} \\ 1 & \cos \frac{1}{2}\gamma & \sin \varphi \end{array} \quad (55)$$

Using  $H_r$  of the particular bearing as given above, projection of (53) over an infinitesimal surface area  $\bar{r}d\bar{\theta}d\bar{s}$  along  $(\bar{i}_{\pm}, \bar{j}_{\pm} = \bar{e}_3 \times \bar{i}_{\pm})$  forms an element of the bearing reaction vector that has radial and tangential components; it can be integrated over the bearing plan form to yield

$$\begin{aligned} \frac{\delta \bar{F}_{r,\pm}}{\pi p_a R^2 (L/D) \varepsilon_r} &= - \int_{\text{transverse span}} \frac{1}{2} \delta P_{\Omega=\pm|\Omega|} H_r \bar{r} d\bar{z} \\ &= -(\bar{i}_{\pm} U_{r,\pm} + \bar{j}_{\pm} V_{r,\pm}) \end{aligned} \quad (56)$$

As written in (56), rotating coordinates  $(\bar{i}_{\pm}, \bar{j}_{\pm})$  are used to describe bearing reaction to circular whirl perturbations. Circular whirl orbits can be viewed in a ground fixed coordinate system with base-vectors  $(\bar{e}_1, \bar{e}_2)$  as

$$\bar{i}_{\pm} = \bar{e}_1 \mp j \bar{e}_2 \quad (57)$$

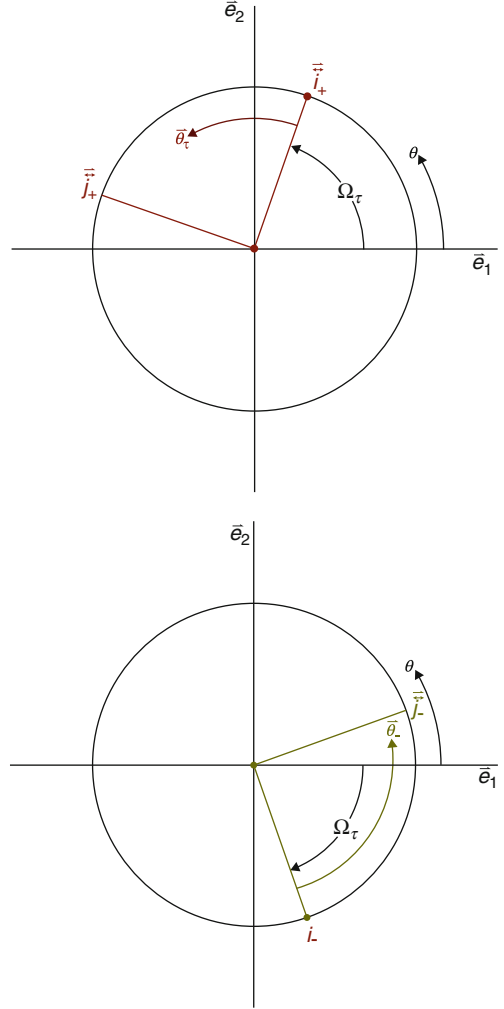
Conversely, a pair of oppositely whirling circular orbits can be linearly combined to form linear oscillations along either ground fixed base-vector.

$$\begin{aligned} \bar{e}_1 \varepsilon_r &= \frac{1}{2}(\bar{i}_{+} + \bar{i}_{-}) \varepsilon_r \\ \bar{e}_2 \varepsilon_r &= \frac{1}{2}(\bar{i}_{+} - \bar{i}_{-}) \varepsilon_r \end{aligned} \quad (58)$$

Cylindrical whirl motion of the rotor is depicted by the trajectories of the rotating base-vectors  $(\bar{i}_{+}, \bar{i}_{-})$  at rate  $\pm |\Omega|$  as illustrated in Fig. 10. The tangential component  $(V_{r,+}, V_{r,-})$  of the perturbation bearing reaction are directed along  $(\bar{j}_{+}, \bar{j}_{-})$ , which are viewed in the ground-fixed coordinate system as

$$\begin{aligned} \bar{j}_{+} &= j \bar{e}_1 + \bar{e}_2 \\ \bar{j}_{-} &= j \bar{e}_1 - \bar{e}_2 \end{aligned} \quad (59)$$

Due to inherent rotational symmetry of the steady-state solution, an examination of  $\bar{e}_1 \varepsilon_r = \frac{1}{2}(\bar{i}_{+} + \bar{i}_{-}) \varepsilon_r$  is sufficient to synthesize the perturbation impedance matrix, the four components computed in (56), namely



Gas Bearings with Narrow Inclined Grooves, Fig. 10  
Cylindrical whirl orbits

$(\bar{i}_{+} U_{r,+}, \bar{j}_{+} V_{r,+}; \bar{i}_{-} U_{r,-}, \bar{j}_{-} V_{r,-})$ , are to be accessed and assembled according to (58) into

$$\delta \bar{F}_1 / \varepsilon_r = -\frac{1}{2} \left[ (\bar{i}_{+} U_{r,+} - \bar{j}_{+} V_{r,+}) + (\bar{i}_{-} U_{r,-} - \bar{j}_{-} V_{r,-}) \right] \quad (60)$$

Note that the algebraic signs of  $(\bar{j}_{+} V_{r,+}, \bar{j}_{-} V_{r,-})$  are reversed from those in (56) to render the right-hand convention for  $\delta \bar{F}_1 / \varepsilon_r$ . Its components along  $-(\bar{e}_1, \bar{e}_2)$

are identified as the complex impedance coefficients in the matrix representation; thus,

$$Z_{11} = \frac{1}{2}(\sum U_r - j\sum V_r), \quad Z_{21} = \frac{1}{2}(-j\Delta U_r - \Delta V_r) \quad (61)$$

Since there is rotational symmetry, one can impose the isotropic property

$$Z_{\perp} \equiv Z_{12} = -Z_{21}, \quad Z_{\parallel} \equiv Z_{22} = Z_{11} \quad (62)$$

So that complete radial perturbation impedance matrix is

$$[Z] = \begin{bmatrix} Z_{\parallel} & Z_{\perp} \\ -Z_{\perp} & Z_{\parallel} \end{bmatrix} \quad (63)$$

**Angular Perturbation.** The angular perturbation problem is concerned with a small transverse angular displacement of the rotor about its mass center. The film thickness perturbation is

$$\delta H = H_{\vartheta} \varepsilon_{\vartheta} \cos \theta = \varepsilon \cos \theta \quad (64)$$

Note that both (47) and (64) result in  $\delta H = \varepsilon$ ; hence, the angular perturbation problem is also governed by (52). However,  $\varepsilon = H_{\vartheta} \varepsilon_{\vartheta}$  is not a constant because an angular displacement brings about film thickness perturbation of the surface point that depends on both its distance from the origin and its orientation.

The angular displacement has components  $(\delta \xi, \delta \eta)$  that are infinitesimal right-hand rotations about the base-vectors  $(\bar{e}_1, \bar{e}_2)$ . Coordinates of a surface point are

$$r(\bar{e}_1 \cos \theta + \bar{e}_2 \sin \theta) + \bar{e}_3 z \quad (65)$$

Its linear displacement caused by the angular displacement is

$$\begin{aligned} \bar{e}_{\vartheta} &= (\bar{e}_1 \delta \xi + \bar{e}_2 \delta \eta) \varepsilon_{\vartheta} \times [r(\bar{e}_1 \cos \theta + \bar{e}_2 \sin \theta) + \bar{e}_3 z] \\ &= (\bar{e}_1 \delta \eta - \bar{e}_2 \delta \xi) z + \bar{e}_3 (\delta \xi \sin \theta - \delta \eta \cos \theta) r \end{aligned} \quad (66)$$

The outward unit normals of various bearings are

Bearing	Outward Unit Normal $\bar{v}$
Journal	$\bar{e}_1 \cos \theta + \bar{e}_2 \sin \theta$
Thrust	$\bar{e}_3$
Conical*	$(\bar{e}_1 \cos \theta + \bar{e}_2 \sin \theta) \cos \frac{1}{2}\gamma - \bar{e}_3 \sin \frac{1}{2}\gamma$
Spherical	$(\bar{e}_1 \cos \theta + \bar{e}_2 \sin \theta) \sin \varphi - \bar{e}_3 \cos \varphi$

(67)

\*Coordinate origin is at apex of cone.

Film thickness perturbation Jacobian is

$$\begin{aligned} H_{\vartheta} &= \bar{v} \cdot (\bar{e}_1 \delta \eta - \bar{e}_2 \delta \xi) z \\ &\quad + \bar{v} \cdot \bar{e}_3 (\delta \xi \sin \theta - \delta \eta \cos \theta) r \end{aligned} \quad (68)$$

For a long journal bearing, the first line in the above formula dominates; whereas for the thrust bearing, only the second line contributes in the perturbation pressure.  $\varepsilon$  would be set as  $\varepsilon = H_{\vartheta} \varepsilon_{\vartheta}$  with  $(\delta \xi, \delta \eta)$  alternatively assigned as  $\varepsilon_{\vartheta}$ . Contrasting features between radial and angular perturbation problems on film thickness perturbation are listed below:

Radial	Angular
In-line harmonic variation	Orthogonal harmonic variation
Independent of other coordinates	Dependent on both radial and axial coordinates

Time-dependence of angular perturbation also can be treated with rotating coordinates. While (57) is still applicable to the rotating base-vectors  $\bar{i}_{\pm}$ , they are no longer the center point of harmonic variation that is the datum for  $\bar{\theta}_{\pm}$ , because the film thickness is orthogonally perturbed. There are two sets of base-vectors; one refers to the angular displacement and is assigned subscript  $\vartheta$  the other refers to the surface point and is assigned subscript  $\sigma$ . Consider the long journal bearing for illustration; the moment arm is mainly the axial coordinate  $z$ , angular displacement takes place about  $\bar{i}_{\vartheta\pm}$  while the resulting surface point displacement is associated with  $-\bar{j}_{\sigma\pm} \equiv \bar{i}_{\vartheta\pm} \times \bar{e}_3$ , which is a quadrant away and is sensitive to  $\text{sgn}\{z\}$ . A depiction of steady-whirl orbits that combine  $\bar{i}_{\vartheta\pm}$  with  $-\bar{j}_{\sigma\pm}$  is shown in Fig. 11.

By comparison of (52) and subsequent moment weighted projected integration over the bearing surface area, it is obtained

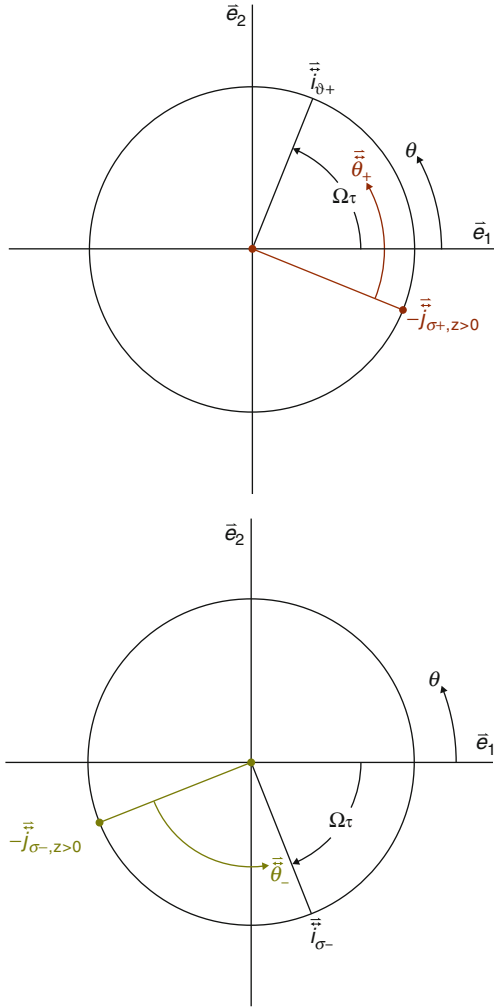
$$\frac{\delta \bar{M}_{\vartheta,\pm} / \varepsilon_{\vartheta}}{\pi p_a R^2 (L/C)(L/D)} = - \left( -\bar{j}_{\sigma\pm} U_{\vartheta,\pm} + \bar{i}_{\sigma\pm} V_{\vartheta,\pm} \right) \quad (69)$$

and

$$Z_{11} = \frac{1}{2} \left( \sum U_{\vartheta} - j \sum V_{\vartheta} \right), \quad Z_{21} = \frac{1}{2} (-j \Delta U_{\vartheta} - \Delta V_{\vartheta}) \quad (70)$$

Again, isotropy is inherent so that

$$Z_{\perp} \equiv Z_{12} = -Z_{21}, \quad Z_{\parallel} \equiv Z_{22} = Z_{11} \quad (71)$$



Gas Bearings with Narrow Inclined Grooves, Fig. 11  
Angular steady-whirl orbits

## Characteristics of Typical Textured Bearings

### $\Lambda$ -Dependence of Textured Gas Bearings

**Limitation of INGA.** Since the isothermal gas film lubrication theory was made known, the ambient pressure is believed to be the proper scale of the film pressure, and because there is an asymptotic limit that governs the achievable film pressure, gas film lubrication has long been regarded as a light-duty application. Whipple's treatment of textured herringbone thrust bearing removes the possibility of an asymptotic limit by imposing the INGA hypothesis:  $\lim_{\delta\theta \rightarrow 0} (\Lambda\Delta\theta) \rightarrow 0$ ; this is reflected

by the last term of (24) so that non-dimensional axial load capacity of a thrust bearing is in Sommerfeld scaling, which is without the limitation of ambient pressure:

$$\frac{\Lambda^{-1}F_z}{\pi R^2 p_a} = \frac{F_z/(\pi R^2)}{6\mu\omega(R/C)^2} \quad (72)$$

The possibility of achieving texture-enhanced film pressure, which can be at a level suitable for applications requiring considerable load capacity, brings forth the need to know the correct upper bound of  $\Lambda\Delta\theta$  for using INGA with confidence. Finite narrow groove analysis (FNGA) was examined by Pan and San Andrés to address this issue; the micro-scale treatment is revised by admitting a non-vanishing  $\lim_{\delta\theta \rightarrow 0} \Lambda\Delta\theta$ .

Sample FNGA results, which are computed for an inward-pumping thrust bearing with texture parameters of Thrust D, shown in Fig. 7 can serve to illustrate the influence of  $\tilde{\Lambda}$ . FNGA pressurization lines are shown in Fig. 12. They are computed for a 2 mm OD bearing running in atmospheric air with  $(C/R) = 10^{-3}$  and textured with 18 grooves at rotation rates up to 2,000,000 rpm. FNGA line for nil  $\tilde{\Lambda}$  degenerates into an INGA line. FNGA lines cannot be distinguished from the INGA line for  $\tilde{\Lambda}$  up to 2.12151, its deviation from the INGA line is quite moderate up to  $\tilde{\Lambda} = 8.48603$  or 200,000 rpm for the 2 mm thrust bearing with  $(C/R) = 10^{-3}$ . The corresponding  $\Lambda = 24.3107$  should be a reasonably safe estimate as upper-bound for INGA usage. For  $\tilde{\Lambda} = 21.2151$  and beyond, both intercepts reduce substantially. There is no observable curvature in all computed FNGA lines.

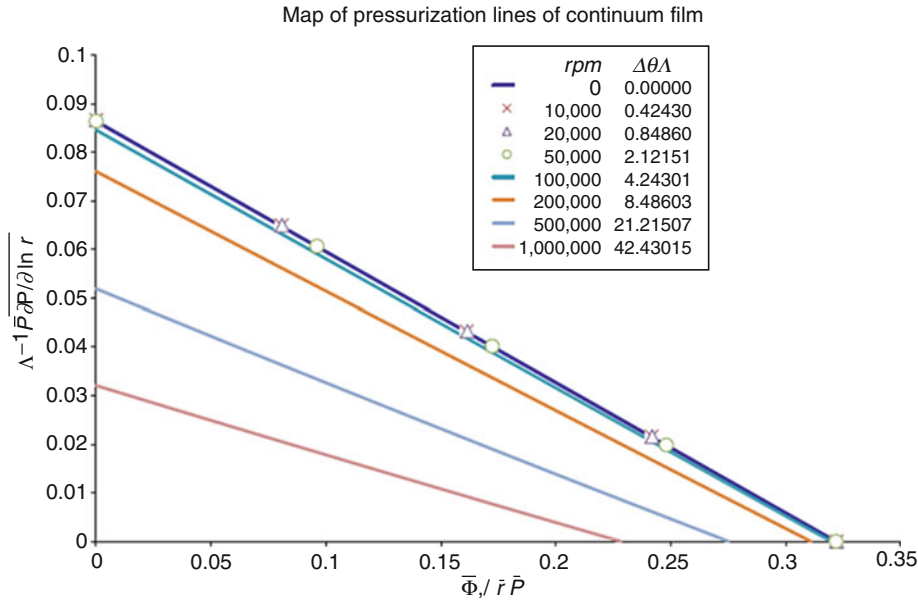
With a finite  $\Delta\theta\Lambda$ ,  $\tilde{p}(\tilde{\theta})$  would not be triangular and its transverse gradient may not be invariant.

They are examined at  $\tilde{\Lambda} = 4.24301$  as illustrated in Fig. 13.

It is seen that departure from INGA features are mainly in the film land as manifested by both the circumferential profile  $(\Delta\theta\Lambda)(\tilde{p} - 1)$  and the transverse pressure gradient  $\Lambda^{-1}\tilde{p}\partial\tilde{p}/\partial(\ln r)$ . The large slope of  $\Lambda^{-1}\tilde{p}\partial\tilde{p}/\partial(\ln r)$  for nil  $\tilde{\Phi}_r$  suggests some reduction of pressurization even though the pressurization line is hardly distinguishable from the INGA line.

**Roles of the Bearing Number.** It has been shown that Sommerfeld scaling applies to the axial load capacity of a thrust bearing. This property is the consequence of (24),  $\Lambda^{-1}(dP_0/\bar{r}d\bar{s})$  is a function of texture parameters only, however, it may not be retained by the stiffness of a textured journal bearing.





Gas Bearings with Narrow Inclined Grooves, Fig. 12 Pressurization line of inward-pumping thrust bearing with finite  $\Delta\theta\Lambda$

The radial perturbation problem of the journal bearing is governed by (52), in which  $\delta P$  is not uniformly multiplied by  $\Lambda^{-1}$  to assure validity of Sommerfeld scaling. Non-dimensional radial perturbation impedance is given by (63). Given a single set of texture parameters, e.g., as listed in the caption of Fig. 7, there are four design variations for the textured journal bearing: combining alternatives of full-herringbone and half-herringbone plan forms with texturing either the stationary bushing surface or the rotating journal. The impedance components for  $L/D = 1.0$  are graphed in Fig. 14. The centered steady-state solution of the full-herringbone bearing operates at the no-flow condition so that the pressure displays complete  $\Lambda$ -independence as demonstrated in Fig. 8. However, all four curves in the top pane of Fig. 14 vary with  $\Lambda$ . All  $K_{21}$  curves tend to approach null value while the corresponding  $K_{11}$  appears to settle down to a lower level. These are symptoms of the asymptotic conditions; thus texturing does not eliminate  $\Lambda$ -dependence in the absolute sense, it does postpone limitation of film pressurization by ambient beyond a relatively large  $\Lambda$ , at which effects of non-vanishing  $\Lambda\Delta\theta$  would require attention.

The  $\Lambda$ -dependency question is especially interesting for the inward-pumping thrust bearings in view of the pressure profiles shown in Fig. 9. A double-acting arrangement is assumed with a pair of identical inward-pumping

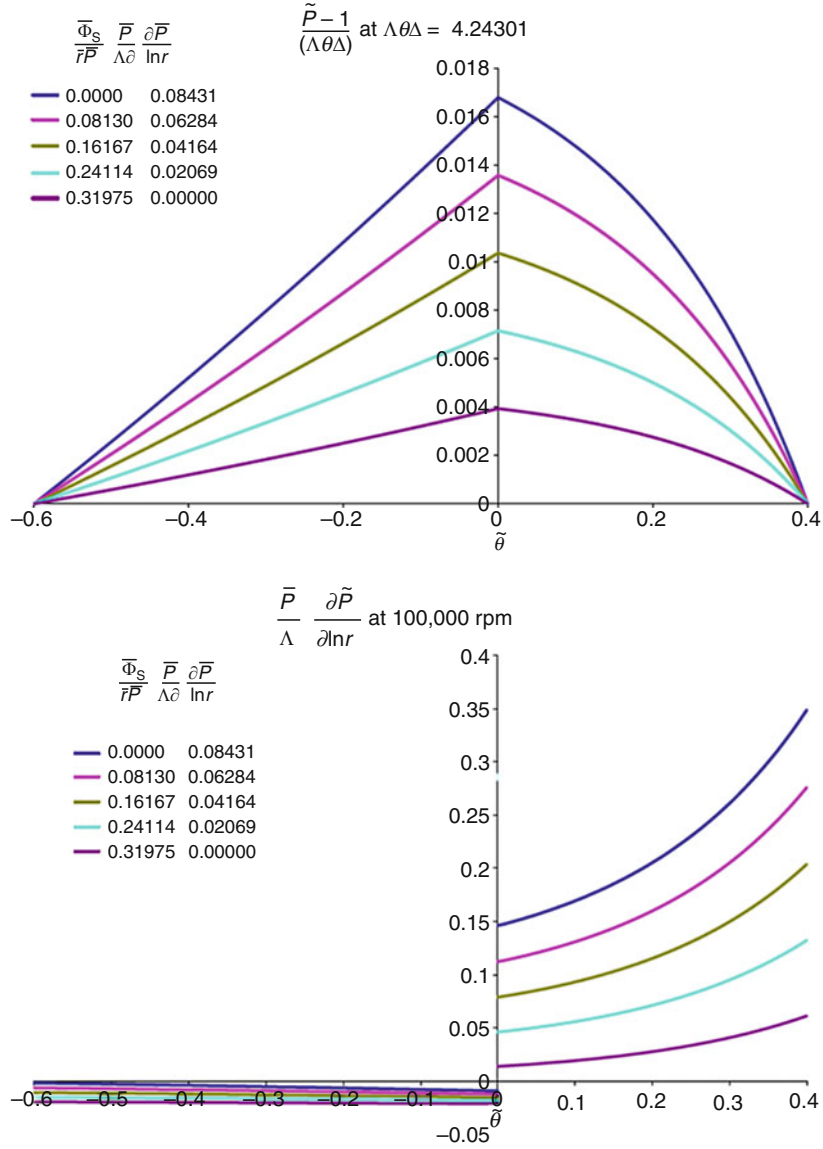
bearings that exert the same axial force as pre-load on each other to maintain a centered equilibrium state. Figure 15 shows the preload and the static stiffness of each bearing in this arrangement.  $\Lambda$ -dependency is practically eliminated altogether.

#### Instabilities of Textured Gas Bearings

Consideration of  $\Lambda$ -dependency must not overlook the possibility of instability. Textured gas bearings are susceptible to instabilities in various DOFs.

**Axial Instability.** The inward-pumping thrust bearing and the spherical bearing can both encounter axial instability upon the occurrence of  $dV_z/d\Omega|_{\Omega=0} < 0$ . Existence of this phenomenon is not unlike the externally pressurized gas bearings (► [Self-Excited Gas Bearing Instabilities](#), ► [Pneumatic Hammer](#)).

Figure 16 presents a graph that superimposes instability parameters, neutral frequency ratio  $\Omega_{\text{neutral}}$  and upper-bound threshold mass  $\mathcal{M}_{\text{threshold}}$  for inward-pumping thrust and spherical bearings. Incipience of potential axial instability occurs at  $\Lambda_{\text{incipience}} = 17.3$  for the thrust bearing, where the static coefficient of damping intercepts the abscissa with  $\Omega_{\text{neutral}} = 0$  and  $\mathcal{M}_{\text{threshold}} \rightarrow \infty$ ; due to the latter values, axial instability tends to evolve slowly. However, as  $\Lambda_{\text{incipience}}$  is exceeded,  $\Omega_{\text{neutral}}$  rises rapidly and  $\mathcal{M}_{\text{threshold}}$  becomes smaller and levels off at  $\Lambda \approx 20$ . The axial static coefficient of damping of the spherical bearing

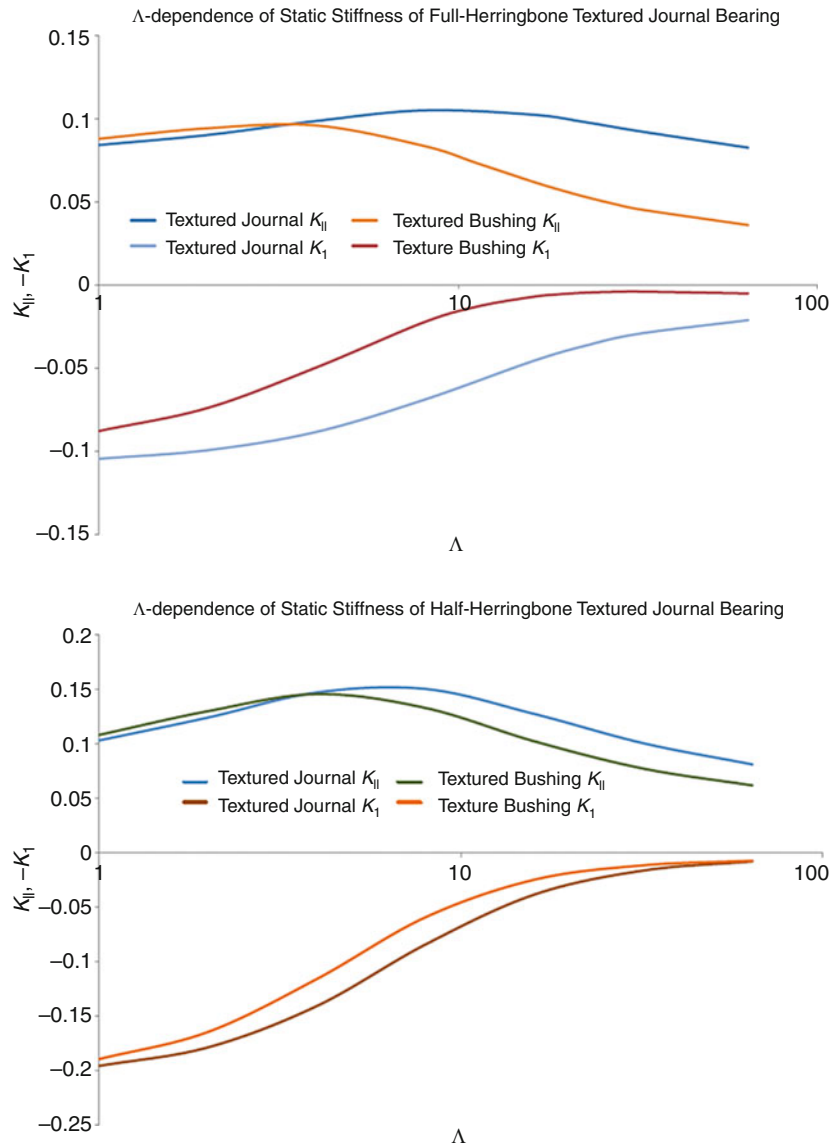


Gas Bearings with Narrow Inclined Grooves, Fig. 13 Micro-scale profiles of  $\tilde{P}$  and its transverse gradient at finite  $\Delta\theta\Lambda$

remains positive until  $\Lambda_{\text{incipience}} = 184$ . The vast difference in values of  $\Lambda_{\text{incipience}}$  is due to the uniform film thickness perturbation in the case of the thrust bearing such that perturbed pressurization in the outer annulus is synchronized with the squeezed film effect, the last term of (39). In contrast, film thickness perturbation in the case of the spherical bearing is mostly confined to the smooth sub-domain of the inner annulus; there is minimal perturbed pressurization in the outer annulus.

*Cylindrical Whirl.* Due to an interaction between the squeeze film and the Couette flux, which share the common factor  $(1 - 2\Omega)$  in (52), there is a reflection property of the perturbation bearing reactions at any operating  $\Lambda$ :

$$\begin{aligned} U_r|_{\text{textured bushing}}(\Omega) &= U_r|_{\text{textured journal}}(1 - \Omega) \\ V_r|_{\text{textured bushing}}(\Omega) &= -V_r|_{\text{textured journal}}(1 - \Omega) \end{aligned} \quad (73)$$



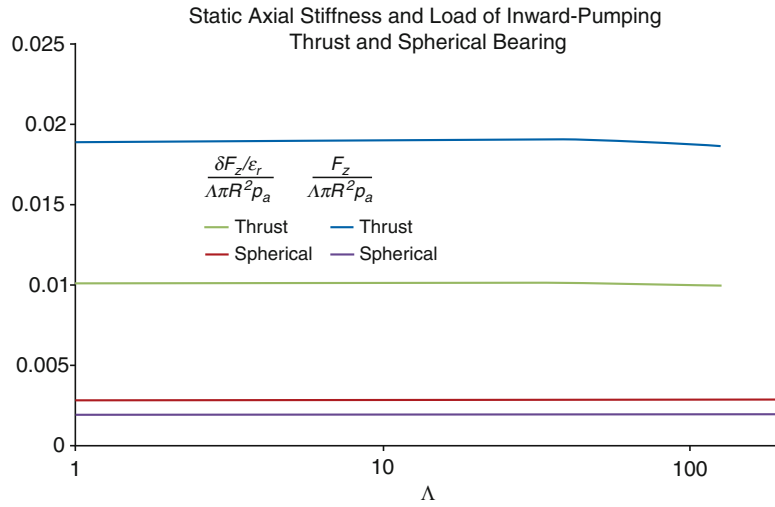
Gas Bearings with Narrow Inclined Grooves, Fig. 14  $\Lambda$ -dependence of static impedances of herringbone journal bearings

Because incipience of potential cylindrical whirl instability depends on a null condition of  $V_r(\Omega_{\text{neutral}})$ , there is a special feature in the whirl instability maps of the four variations of herringbone journal bearings that share the same texture parameters as illustrated in Fig. 17.

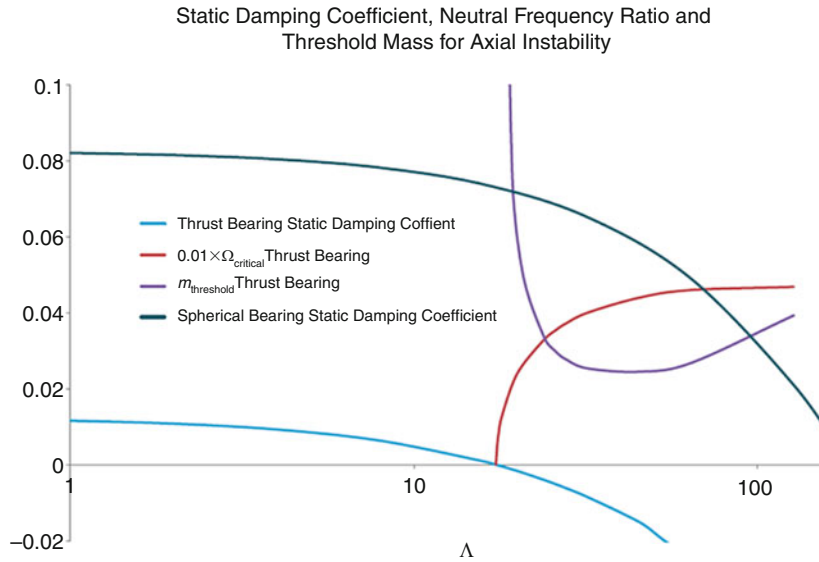
The mirror property stated by (73) is revealed by the symmetrical arrangement of  $\Omega_{\text{neutral}}$  on either side of the chain line that marks  $\Omega = 0.5$ ; those with textures on the journal are above and those with textures on the

bushing are below. The lowest value of  $\Omega_{\text{neutral}}$  is that of the full-herringbone bearing textured on the bushing at about  $\Lambda = 32$ ; the corresponding  $M_{\text{threshold}}$  peaks to about 63. There is clearly the possibility to select a design for the largest possible whirl stability margin.

**Conical Whirl.** For the angular **DOF** concerning conical whirl instability, the precession d'Alembert moment due to the polar moment of inertia of the rotor enters into the moment equilibrium condition Self-Excited Gas



Gas Bearings with Narrow Inclined Grooves, Fig. 15 Axial stiffness and load of inward-pumping thrust and spherical bearings



Gas Bearings with Narrow Inclined Grooves, Fig. 16 Static axial damping coefficient and stability parameters of inward-pumping bearings

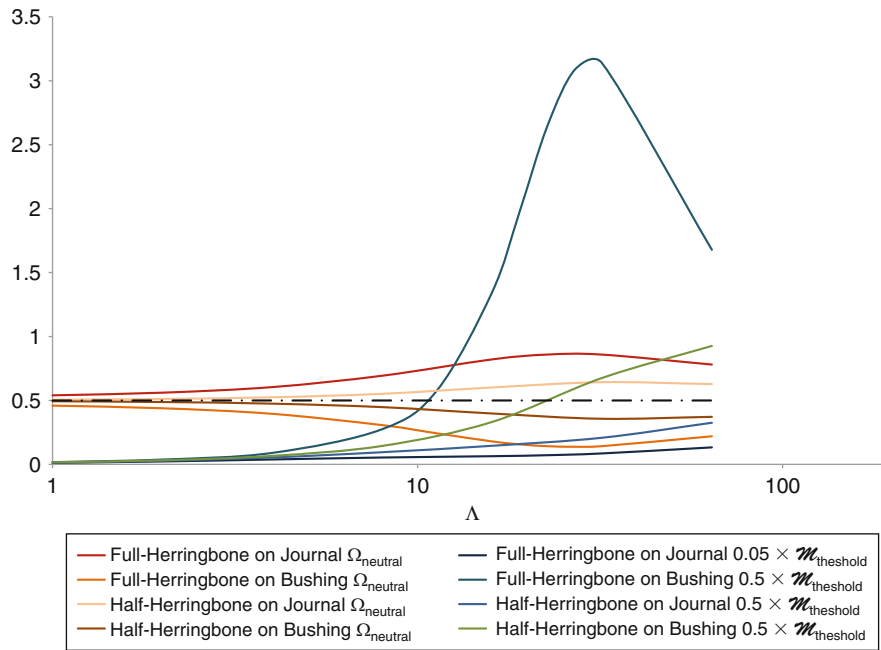
Bearing. Vanishing of  $V_{\theta}(\Omega_{\text{incipience}})$  is still a requirement for the incipience of potential conical whirl instability, the neutral state of conical whirl instability is

$$J_{T,\text{threshold}} = \frac{(\Omega_{\text{incipience}} \omega)^{-2} U(\Omega_{\text{incipience}})}{[1 \mp \Omega_{\text{incipience}}^{-1} (J_P / J_T)]} > 0 \quad (74)$$

If the above positivity requirement is satisfied, a state of neutral stability is identified at  $\Omega_{\text{neutral}} = \Omega_{\text{incipience}}$  then (74) renders the threshold value of the transverse moment of inertia for conical whirl instability. Unique in the conical whirl instability is that the two parameters for the state of neutral instability, namely  $(\Omega_{\text{neutral}}, J_{T,\text{threshold}})$ , are filtered by the precession d'Alembert moment.



**Whirl Frequency Ratio and Threshold Mass for Whirl Instability of Herringbone Journal Bearings**



Gas Bearings with Narrow Inclined Grooves, Fig. 17 Whirl stability map of herringbone journal bearings

## Key Applications

Petrol chemical processing equipment  
Machine tools  
Computer hard disc drive spindle  
Inertia navigation instruments  
MEMS

## Nomenclature

Roman letters	
$c_{\theta, \Gamma}$ , etc.	Texture coefficients in formula for $\overline{\Phi}_{\theta}$ , (18)
$c_{s, \Gamma}$ , etc.	Texture coefficients in formula for $\overline{\Phi}_{\theta}$ , (19)
$C$	Nominal clearance, $m$
$D$	Diameter of journal bearing, overall diameter of bearing cartridge, $m$
$\vec{e}_1, \vec{e}_2, \vec{e}_3$	Ground fixed base-vectors
$h$	Film thickness, $m$
$H$	Normalized film thickness, $= h/C$
$H_z, H_r, H_{\vartheta}$	Film thickness perturbation Jacobians of the respective <b>DOF</b>
$\overline{H^3}$	$\equiv H_{\text{groove}}^3 H_{\text{land}}^3$

$\vec{i}, \vec{j}$	Circumferential and transverse base-vectors
$j$	$\sqrt{-1}$
$k$	Index of computation grid point
$K_{11}, K_{21}$ , etc.	Components of static stiffness matrix
$\ell$	Axial offset of bearing center from apex of cone, $m$
$L$	Overall bearing length, $m$
$L, L^*, L'$	Lengthwise dimensions of conical bearing cartridge, $m$
$M_{\text{threshold}}$	Non-dimensional mass parameter that would trigger instability
$p$	Film pressure, $MPa$
$\bar{p}$	Cross-pattern averaged film pressure, $MPa$
$\tilde{p}$	$\equiv \bar{\Lambda}^{-1}(\bar{p} - 1)$
$\delta \tilde{p}$	Amplitude of micro-scale cross-pattern film pressure variation, $= \tilde{p} _{\vartheta=0} - \tilde{p} _{\vartheta=-\alpha}$
$p_a$	Ambient pressure, $MPa$
$P$	Macro-scale normalized film pressure, $\equiv p/p_a$
$\delta P$	Perturbation film pressure, $= (u + jv)\varepsilon$
$\bar{P}$	Macro-scale normalized averaged film pressure, $\equiv \bar{p}/p_a$

$\bar{p}$	Micro-scale normalized film pressure, $\equiv p/\bar{p}$
$r$	Cylindrical radius, $m$
$r_i$	Inside radius of thrust bearing, $m$
$r_m$	Medium radius of no through-flow herringbone thrust bearing, $m$
$\bar{r}$	Normalized cylindrical radius, $\equiv r/R$
$R$	Outside radius of thrust bearing, reference cylindrical radius, $m$
$s$	Generic transverse coordinate, $m$
$s_m$	Medium generic transverse coordinate of no through-flow herringbone plan form, $m$
$\bar{s}$	Generic normalized transverse coordinate, $\equiv s/R$
$\tilde{s}$	Micro-scale transverse coordinate
$t$	Time, $s$
$u, v$	Real and imaginary part of $\delta p$
$U_z, V_z$	Real and imaginary part of $(\pi R^2 p_a)^{-1} (\delta F_z / \hat{e}_z)$ , (45)
$U_{r,\pm}, V_{r,\pm}$	See (56), components of $-\left[\pi p_a R^2 (L/D)\right]^{-1} (\delta \vec{F}_{r,\pm} / \hat{e}_r)$
$z$	Axial distance from mid-plane of bearing, $m$
$\bar{z}$	$\equiv z/R$
$Z_{11}, Z_{21}$ , etc.	Components of impedance matrix
<b>Greek letters</b>	
$\alpha$	Groove width fraction
$\beta$	Inclination of groove texture measured from $\vec{i}_r$ , ° or <i>radian</i>
$\varepsilon$	Infinitesimal non-dimensional displacement
$\varphi$	Co-latitude angle of sphere, <i>radian</i>
$\vec{\Phi}$	Normalized flux vector, $= \vec{i} \Phi_\theta + \vec{j} \Phi_s$
$\vec{\Phi}_\theta, \vec{\Phi}_s$	Locally scaled circumferential and transverse fluxes
$\gamma$	Half apex angle, <i>radian</i>
$\Gamma$	Normalized groove depth, $= \delta/C$
$\vec{\lambda}, \vec{v}$	Base-vectors along longitudinal and normal directions of texture
$\Lambda$	Global bearing number $= 6\mu(\omega/p_a)(R/C)^2$
$\tilde{\Lambda}$	Micro-scale local bearing number, $\equiv \Lambda \Delta\theta (r/R)^2 (p_a/\bar{p})$
$\theta$	Circumferential angular coordinate, <i>radian</i>
$\Delta\theta$	Texture period, <i>radian</i>
$\bar{\theta}$	Micro-scale circumferential coordinate
$\tilde{\theta}$	Circumferential coordinate in the rotating coordinate system, <i>radian</i>
$\tau$	Normalized time, $\omega t$

$\omega$	Shaft rotational rate, <i>rad/s</i>
$\Omega$	Normalized oscillation rate, $= 2\pi f / (\frac{1}{2}\omega)$
$\Omega_{\text{neutral}}$	Value of $\Omega$ at incipience of neutral instability
$\Psi$	Through-flow invariant, (24)
<b>Subscripts</b>	
0	Pertaining to the steady-state macro-scale solution
$\pm$	Pertaining to $r$ or $\vartheta$ perturbation computation with $\Omega = \pm \Omega $ in the rotating coordinate system
Groove, land	Pertaining to the respective part of the texture pattern
$m$	Internal boundary between sub-domain annuli
$z, r, \vartheta$	Pertaining to axial, radial and angular degrees of freedom of the supported rotor
$\sigma$	Pertaining to the surface point in the angular perturbation problem
<b>Mathematical symbols</b>	
$\vec{\nabla}$	Normalized gradient operator in micro-scale, $\equiv r(\Delta\theta)^{-1} \text{grad}$
$\vec{\nabla}$	Normalized gradient operator, $\equiv R \text{grad}$
$\Delta(\ )$	$\equiv (\ )_{\text{groove}} - (\ )_{\text{land}}$
$\overline{(\ )}$	$\int_{-\alpha}^{1-\alpha} (\ ) d\bar{\theta}$ or $\int_{-\alpha}^{1-\alpha} (\ ) d(\bar{s}/\tan \beta)$

## Cross-References

### ► Self-Excited Gas Bearing Instabilities

## References

- D. Bonneau, J. Huitric, B. Tournier, Finite element analysis of grooved gas thrust bearings and grooved gas face seals. *ASME J. Tribol.* **115**, 348–354 (1993)
- W.J. Harrison, The hydrodynamic theory of lubrication with special emphasis to air as a lubricant. *Trans. Camb. Phil. Soc.* **xxii**, 373–388 (1913)
- C.H.T. Pan, L. San Andrés, The narrow groove analysis revisited, in *Proceedings of WTC2005, World Tribology Congress III*, Paper No. WTC2005-63803, Washington, D.C., USA, 12–16 Sept 2005
- M.M. Reddi, Y.T. Chu, Finite element solution of the compressible lubrication problem. *ASME J. Lubrication Technol.* **92**, 495–503 (1970)
- J.H. Vohr, C.H.T. Pan, Gas-lubricated spin-axis bearings for gyroscopes, Technical Report MTI-68TR29, prepared for Office of Naval Research under Contract N00014-67-C-0530 NR 062-370/2-21-27 (1968)
- R.T.P. Whipple, Theory of the spiral grooved thrust bearing with liquid or gas lubricant, Atomic Energy Research Establishment Technical Report No. 622 (1951)
- N. Zirkelback, L. San Andrés, Effect of frequency excitation on the force coefficients of spiral groove thrust bearings and face gas seals. *ASME J. Tribol.* **121**(4), 853–863 (1999)

## Gas Laby Seal

### ► Labyrinth Seal

## Gas Lubricated Seal

### ► Mechanical Seals

## Gas Lubrication Applications

### ► Gas Bearing Applications

## Gas Nitriding

TOMASZ BABUL

Heat Treatment Department, Institute of Precision Mechanics, Warsaw, Poland

### Synonyms

Ammonia nitriding; Nitriding; Nitrogen hardening

### Definition

Nitriding is a thermo-chemical treatment process consisting of surface impregnation of metals and their alloys – mainly ferrous – with nitrogen. This process is most often conducted at temperatures within the range of 500–600°C in a gaseous atmosphere generated from partially dissociated ammonia.

### Scientific Fundamentals

#### The Iron–Nitrogen Equilibrium System

Solubility of molecular nitrogen does not practically occur in iron in its liquid state and, at a temperature of 1,600°C, amounts to only 0.045%. For this reason the process is conducted in an atmosphere in which nitrogen occurs in the atomic state.

The concentration of nitrogen in iron corresponds to the Fe–N phase diagram (Fig. 1). The following phases can be distinguished:

- $\text{Fe}\gamma\text{N}$  – solid solution of nitrogen in Fe with an A2 body-centered cubic lattice and maximum concentration of nitrogen of 0.11% at the eutectoid temperature of 590°C.

- $\text{Fe}\gamma'\text{N}$  – solid solution of nitrogen in  $\text{Fe}\gamma'$  with an A1 face-centered cubic lattice and maximum nitrogen concentration of 2.8% at a temperature of 650°C.
- The  $\gamma'$  phase (the  $\text{Fe}_4\text{N}_{1-x}$  nitride) solid solution of nitrogen in a matrix of  $\text{Fe}_4\text{N}$  with an A1 face-centered cubic lattice. Above 680°C this phase transforms to the  $\varepsilon$  –  $\text{Fe}_2\text{N}_{1-x}$  phase.
- The  $\varepsilon$  phase ( $\text{Fe}_2\text{N}_{1-x}$  nitride) solid solution in a matrix of the  $\text{Fe}_2\text{N}$  nitride with a hexagonal A3 lattice.
- The  $\xi$  phase –  $\text{Fe}_2\text{N}$  nitrided with a rhombic lattice, stable up to 450°C.

Other phases formed during processes accompanying the thermal treatment of nitrided cases are the following:

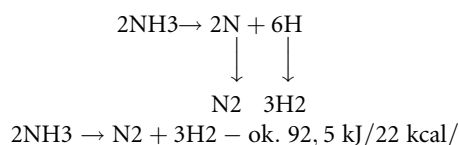
- The  $\text{Fe}\alpha'_\text{N}$  – (nitromartensite) phase formed in conditions of fast cooling of nitroaustenite  $\text{Fe}\gamma\text{N}$ , resulting in the supersaturated  $\text{Fe}\alpha'_\text{N}$  phase, with a body-centered tetragonal lattice.
- The  $\text{Fe}\alpha''_\text{N}$  phase – metastable  $\text{Fe}_{16}\text{N}_2$  nitride with a body-centered tetragonal lattice. It is formed during the process of tempering of the nitromartensite  $\text{Fe}\alpha'_\text{N}$  at temperatures up to 180°C, precipitating out of the  $\varepsilon$  phase. Above 250°C it is transformed into the  $\text{Fe}_4\text{N}_{1-x}$  phase.

### Mechanism of Formation of Nitride Phases

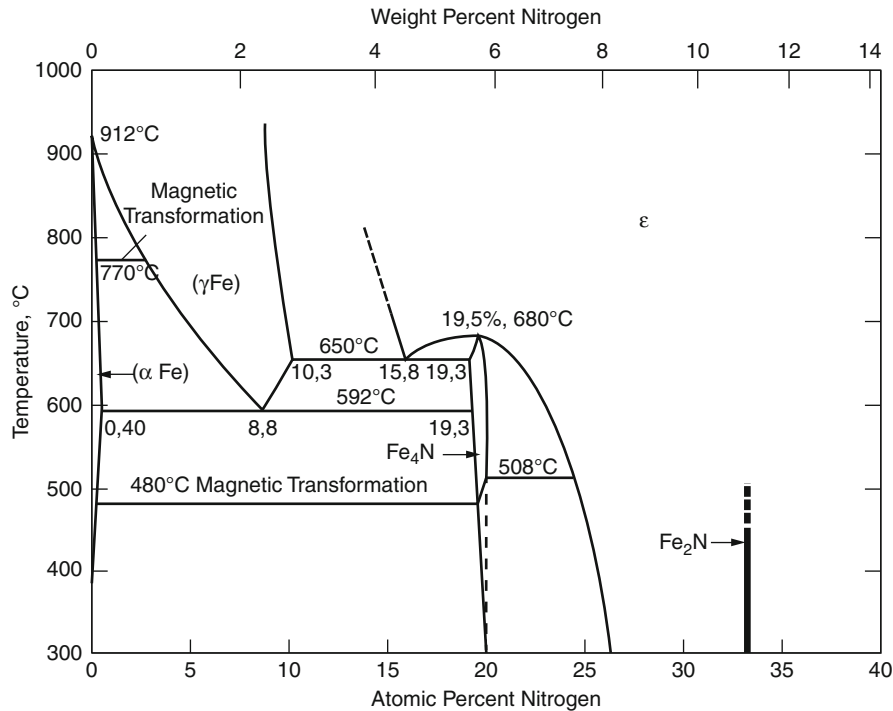
Three basic steps of nitriding are distinguished, dissociation of ammonia, sorption, and diffusion, which are decisive to the concentration of nitrogen in the layer, as well as to the duration of the process.

### Dissociation

Dissociation of ammonia is affected by temperature. Dissociation of ammonia under normal conditions takes place at a temperature of 900°C, but when a catalyst (e.g., iron) is introduced, it begins at 300°C. Development of nitrided surfaces promotes the process of dissociation. During ammonia dissociation, active atomic nitrogen is formed (*in statu nascendi*), quickly recombining back to the molecular state:



The time of existence of nitrogen in the atomic state is 1–1.5 s.



Gas Nitriding, Fig. 1 The Fe-N Phase Diagram (Massalski 1990)

Thermodynamic activity of nitrogen in iron ( $a_N$ ) is expressed by the correlation characterizing partial dissociation of ammonia:

$$a_N = k_1 \frac{P_{NH_3}^{3/2}}{P_{H_2}}$$

where:

$P$  = partial pressure  $NH_3$  and  $H_2$

$k_1$  = equilibrium constant.

The ratio of partial pressure of dissociated ammonia to that of hydrogen  $P_{NH_3}/P_{H_2}^{3/2}$  is accepted as the nitriding potential of the atmosphere ( $N_p$ ). The rise of partial pressure (amount) of hydrogen and drop of the partial pressure (amount) of non-dissociated ammonia in the gaseous phase causes a lowering of its nitriding ability.

Dissociation of ammonia is favored by a developed surface to be nitrided. A drop in the surface roughness parameters, such as  $R_a$  and  $R_z$ , causes the retardation of the nitriding process.

### Sorption of Nitrogen

Sorption comprises two processes: *adsorption*, which is the sedimentation and self-location of atoms and particles of

nitrogen on the surface of the nitrided metal, and *absorption* of the adsorbed particles by the metal.

Sorption processes are affected by smoothness, contaminations, and flaws, as well as chemical changes at the surface of the metal. Atoms of nitrogen first come into contact with the surface of the metal. Next, as the process continues, they come into contact with nitride phases created at the surface, with a microstructure and chemical composition different from those of the metal. The basis of adsorption processes during nitriding is the mutual electron reaction between atoms of nitrogen with atoms of the metal.

The process of diffusion of nitrogen into the metal causes the lowering of concentration of the nitrogen at the surface of the metal and progresses properly when the rate of adsorption is somewhat higher than the rate of diffusion ( $v_s \geq v_D$ ). Contemplating the processes of dissociation, gas nitriding, it is necessary to create such conditions that:

$$v_d \geq v_s \geq v_D$$

The condition of diffusion of nitrogen into  $\alpha$  and  $\gamma$  iron is its higher chemical potential in the gas phase than that in the solid phase. The difference in potentials  $\Delta_F = F_{gas}^N - F_{Fe}^N$  is the thermodynamic driving force of

the nitriding process. The rate of permeation of nitrogen into the metal is characterized by a gradient of chemical potentials and a diffusion coefficient.

### Internal Nitriding

Internal nitriding is the term given to nitriding processes of metal alloys in which layers are formed, composed of a solid solution  $\alpha$  (N) with precipitated nitride phases of alloying additives and a surface layer, composed of metal nitrides.

*Internal nitriding of the first kind* is used to describe nitriding of metal alloys with, for example, alloying additives that have a lower chemical affinity to nitrogen in comparison with the nitrided alloy, such as alloys of iron with Co, Ni, and Cu additives.

*Internal nitriding of the second kind* is used to describe nitriding of metal alloys with alloying additives that have a greater chemical affinity to nitrogen in comparison with the affinity of the nitrided alloy, such as alloys of iron with Mo, Cr, Mn, and V additives (Lachtin and Kogan 1976).

### Nitriding Kinetics

During the process of nitriding, several phases are formed:  $\xi$  –  $\text{Fe}_2\text{N}$ ,  $\varepsilon$  –  $\text{Fe}_2\text{N}_{1-x}$ , or  $\gamma'$  –  $\text{Fe}_4\text{N}$ , which all hinder the growth by diffusion of the nitrided layer. This stems from the difference in diffusion coefficients of nitrogen through these phases. The diffusion coefficients in nitrided phases are as follows;

$$D_N \varepsilon = \frac{1}{60} D_{N'}^{\alpha}, \quad D_N^{\gamma'} = \frac{1}{25} D_{N'}^{\alpha},$$

$$\text{while } D_N^{\alpha} \approx \frac{1}{50} D_N^{\alpha},$$

Alloying additives in nitriding steels, such as Cr, Mo, V, Ti, and others, raise the activity of nitrogen in ferrite and at the same time lower the coefficient of diffusion of nitrogen. These values have not yet been calculated. For ferrous alloys containing strong nitride-forming elements, such as titanium, chromium, and vanadium, the depth of the zone of internal nitriding ( $x$ ) is calculated from the formula:

$$x^2 \approx \frac{2N(\%at)}{rM(\%at)} D\tau$$

where:

$r$  – proportion of concentration of N atoms at the surface of the solution layer (may be determined by the ratio of  $\text{NH}_3$  to  $\text{H}_2$  in the gaseous atmosphere) to the concentration of atoms of alloying additives M in the alloy prior to nitriding.

$D$  – coefficient of diffusion of nitrogen

$\tau$  – duration of process

**Gas Nitriding, Table 1** Criteria classification of gas nitriding process

Criterion	Type of process
Microstructure of nitrided material	Ferritic (<590°C)
	Austenitic (>590°C)
Time	Long duration (one-, two-, and three-stage);
	Short duration
Method of layer formation	Conventional (not controlled)
	Controlled

An effect of the diffusion of nitrogen into the core of the metal is the phenomenon of dimensional growth of the treated component. It is caused mainly by the formation of the nitride layer at the surface. It is accepted that for carbon steels two-thirds of the thickness of the nitride layer is within the nitrided metal, while one-third of the thickness of the nitride layer is above the original surface (Zyśk 2008).

### Types of Gas Nitriding

Gas nitriding can be classified according to several criteria given in Table 1.

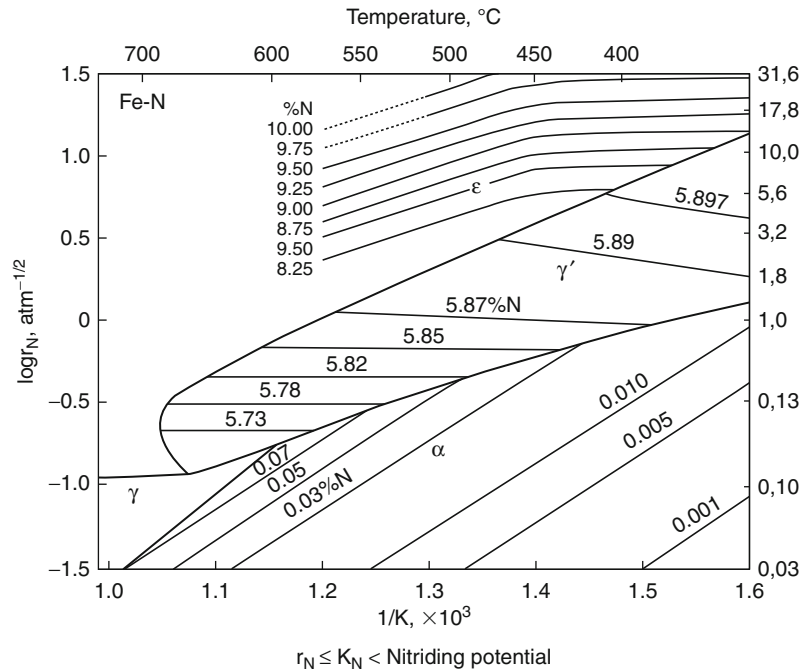
*Conventional nitriding*, used to this day, conducted based on measurements of the degree of ammonia dissociation, is being replaced ever more frequently by *controlled nitriding*, based on the measurement of the nitriding potential.

The process of controlled nitriding is conducted, based on correlations, determined by E. Lehrer, between the nitriding potential of the atmosphere and the temperature of the process (see Fig. 2) and the iron nitride phases (i.e.,  $\alpha$ ,  $\gamma$ ,  $\gamma'$ ,  $\varepsilon$ ) being formed. From the diagram it follows that with the rise of temperature, the nitriding potential of the atmosphere should be reduced in order to maintain border values of phases  $\alpha/\gamma'$  and  $\gamma'/\varepsilon$  in equilibrium with each other.

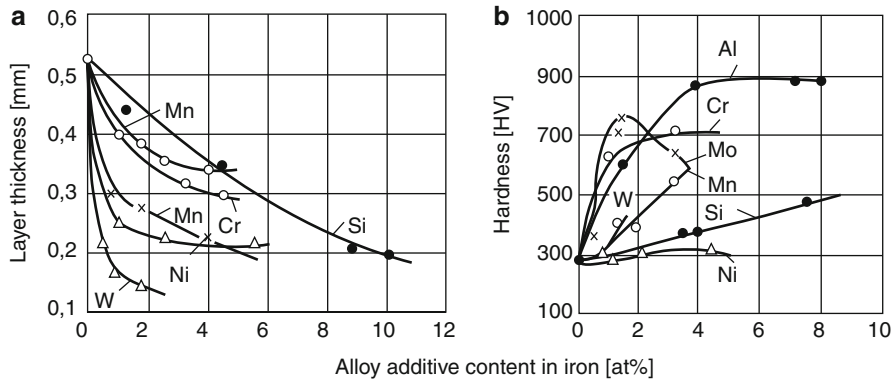
### Effect of Alloying Elements

Aluminum, vanadium, chromium, and molybdenum are the elements most frequently added to steels designated for nitriding, on account of their greatest affinity to nitrogen. Transition metals lower the activity of nitrogen in iron and raise its distribution in iron. This causes the lowering of the coefficient of diffusion of nitrogen, and by the same token reduces the thickness of layers, while at the same time raising hardness. See Fig. 3 (Lakhtin).





**Gas Nitriding, Fig. 2** The E. Lehrer diagram. Zones of nitrided phases in equilibrium with the nitriding atmosphere (Lehrer (1930) modified by L. Małdziński)



**Gas Nitriding, Fig. 3** Effect of alloying elements in iron on the thickness of the nitride layer (a) and on hardness (b). Nitriding at 550°C for 24 h (Lachtin and Kogan 1976)

### Prior Heat Treatment

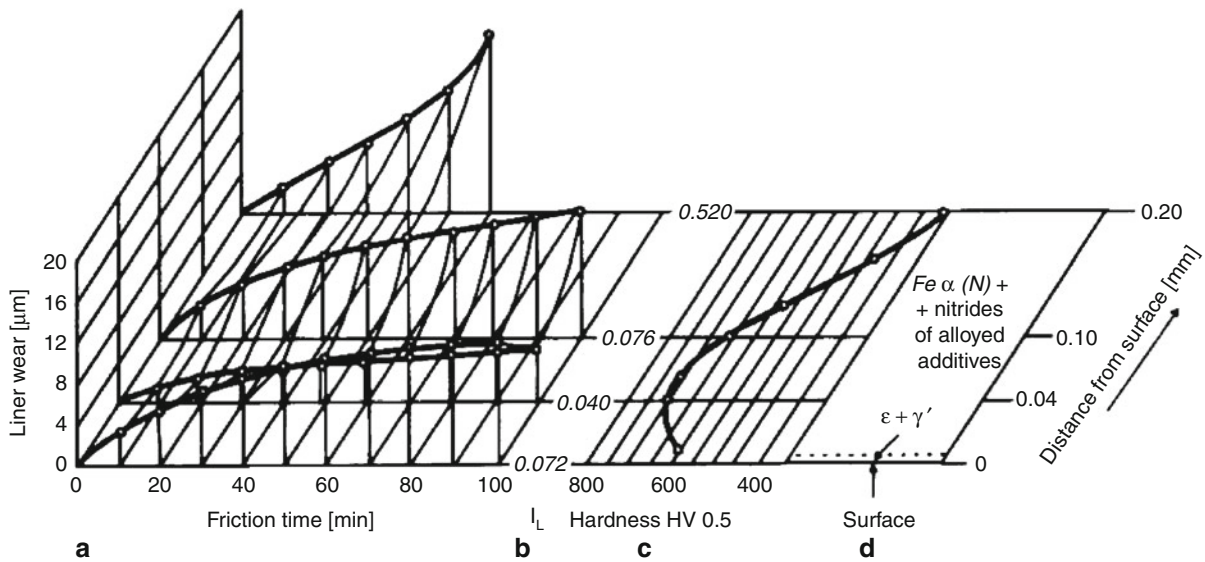
Before nitriding, all hardenable steels must be hardened and tempered. Usually the minimum tempering temperature should be about 20°C higher than maximum nitriding temperature because the tempering temperature must guarantee the stability of structure obtained in nitriding process.

### Nitriding Steels

#### Prior Heat Treatment

Nitriding steels can be classified as follows:

1. Plain carbon steel (e.g., 1,010, 1,020, 1,045)
2. Alloyed steels containing aluminum (e.g., 41CrAlMo7)



**Gas Nitriding, Fig. 4** Tribological characteristic of the nitrided layer: (a) linear wear as a function of the wear path and distance from surface; (b) intensity of wear as a function of distance from surface; (c) hardness profile in the layer; (d) layer microstructure (Burakowski et al. 1986)

- Low alloy steels, chromium steels, and medium carbon steels belonging to the group (SAE 4,100, 4,300, 5,100, 6,100, 8,600, 8,700 as well as 9,800) (► [Gear Materials](#))
- Hot work tool steels containing 5% chromium (e.g. H11, H12, and H13)
- Low alloy chromium and low carbon steels from the groups 3,300, 8,600, and 9,300
- Air hardening tool steels such as A-2, A-6, D-2, D-3, and S-7
- High-speed steels, such as M1 and M4
- Low carbon stainless steels (Armco), such as 30, 40, and 60
- Ferritic and martensitic stainless steels from the 400 and 500 groups
- Austenitic stainless steels from the 200 and 300 groups
- Precipitation hardening stainless steels, such as 13-8 PH, 15-5 PH, 17-4 PH, 17-7 PH, A-286, AM350 and AM355.

## Key Application

The nitriding process is conducted, as a rule, after quenching and tempering, the object of which obtains appropriate mechanical properties in the cross-section, avoiding of excessive deformation after nitriding and obtaining the better properties of the nitrided layer itself.

Nitriding processes are conducted in order to enhance mechanical properties (wear Resistance; see [Fig. 4](#), and fatigue strength [► [Fatigue](#); ► [Fatigue Limit](#)]), as well as corrosion resistance [► [Corrosive Wear](#)] of components applied in various branches of industry; aerospace (shafts, cams, piston rings), energy gears (► [Gear Surface Treatment](#); ► [Wear in Gears](#)), valve stems, machine tooling (spindles, worm gears), automotive (crankshafts, pins, cylindrical tubes), as well as in the textile, railroad, tooling industries, and others.

## Cross-References

- [Corrosive Wear](#)
- [Fatigue](#)
- [Fatigue Limit](#)
- [Gear Materials](#)
- [Gear Surface Treatment](#)
- [Wear in Gears](#)

## References

- T. Burakowski, J. Tacikowski, J. Senatorski, Tribological properties of diffusion layers on steels, in *Advances in Surfaces Treatments Technology. Application Effect*, vol. 2 (Pergamon Press, Heading Hill Hall, Oxford, 1986), pp. 149–159
- J.U.M. Lachtin, J.A.D. Kogan, Azorirovanie stali. (Mašinostroenie, Moskva, 1976)
- E. Lehrer, Über das Eisen-Wasserstoff-Ammoniak-Gleichgewicht. *Z. Elektrochem.* **6**(36), 383–392 (1930)

T.B. Massalski, *Binary Alloy Phase Diagrams* (ASM International, Materials Park, Ohio, 1990)

J. Zyśk, *Rozwój Azotowania Gazowego Stopów Żelaza* (Instytut Mechaniki Precyzyjnej, Warszawa, 2008)

## Gas Nitriding Process

### ► Nitriding and Nitrocarburizing

## Gas Phase Lubrication

### ► Vapor Phase Lubrication for Micro-Machines

## Gas Turbine Engine Lubricants

DAVID E. CHASAN

BASF Corporation, Tarrytown, NY, USA

### Synonyms

Combined cycle turbine lubricant; High temperature lubricants

### Definition

Gas turbines are rotary machines in which air is compressed and combusted with gas or oil to produce high-pressure, high-velocity gas, which, in turn, drives a turbine to produce mechanical power. This power can be used to drive another compressor or a power generator. Steam turbines function similarly, converting the thermal energy of steam into kinetic or rotational energy, which is, in turn, converted into mechanical work. Other turbines employing cogeneration, or combined cycle operation, are increasingly being used. They significantly boost the thermal efficiency of a gas turbine by capturing the exhaust gas for either useful heating or to power a steam turbine. Due to their unique operating conditions, gas turbines require specially designed lubricants to facilitate their operation.

### Scientific Fundamentals

#### Gas Turbines

Gas turbines can vary in size from microturbines to room-sized power generators. They have been used with varying

degrees of success to power railroad locomotives, jet engines, passenger cars, race cars, busses, tanks, and ships as well as to drive pumps and compressors. However, the majority of gas turbines are now used in industrial power generation systems. They range in size from 30 kW to 250 MW and range in thermal efficiency from ~40% for simple to ~60% for combined-cycle turbines.

### Lubrication

Lubricants for turbines are required to form a fluid film between moving parts that will:

- Control friction and wear of surfaces
- Protect against rust and corrosion of precision parts (eliminate water)
- Protect against the accumulation of deposits (sludge and varnish) in the system
- Resist aeration and foaming, which can cause malfunctions
- Act as a coolant to remove heat generated from within or outside the equipment
- Carry away contaminants from internal or external sources
- Often act as a hydraulic medium
- Maintain functionality and longevity at both high and low temperatures
- Seal between moving parts and gasses

### Key Applications

Lubrication requirements for gas and steam turbines are similar but not identical. Protection from exposure to very high temperatures is the dominant concern for gas turbine lubricants. Steam turbines generally have lower heat exposure; however, they encounter significant amounts of water (steam). Turbine fluids may be formulated to perform in either gas or steam turbines or in combined cycle turbines, where a single lubricant circulating system is used to supply both. The properties of such a formulation must meet the requirements of both but can simplify lubrication logistics.

Often the lubricant is supplied to the gas turbine, load equipment, accessories, and controls from a common lubricating system. The lubricant must be supplied at the proper pressure, flow rate, temperature, and cleanliness level for start-up, running, shut down, and cool down of the gas turbine. Since the rotor and housing remain hot for several hours after the gas turbine shuts down, the lubricant must continue to circulate through the bearings to protect Babbitt bearing alloys from overheating. The lubrication system includes a reservoir, pumps, cooler, full-flow filtration, pressure and temperature controls,

Gas Turbine Engine Lubricants, Table 1 Gas turbine lubricant specifications

General electric	Siemens westinghouse	Solar	Alstom	Cincinnati machine	Mitsubishi heavy industries (MHI)
GEK-32568 F	M 55125Z3	ES9-224, Class II	HTG 90117	P-38	MS04-MA-CL001
GEK-101941A	TLV 9013 04				MS04-MA-CL002
GEK-28143A					MS04-MA-CL003
GEK-90117					MS04-MA-CL005
GEK-107395A					

and necessary piping to circulate the lubricant to and from the various components (Smith 1987). The size of the reservoir can vary from 1,000 to 20,000 gal – providing incentive for use of long-life lubricating oils (Hannon 2001).

Recent gas turbine design improvements to increase efficiency and output have put more stress on the gas turbine lubricant. Higher firing temperatures (increased “hot spots”), bearing temperatures (well over 300°C), reduced oil reservoir size, increased peak or cycle running, and use of common reservoirs (combining bearing oil with control oil) have necessitated upgrades in lubricant quality. Intermittent and cyclic operation (as opposed to continuous run) has been directly linked to problems such as the formation of increased insoluble degradation products that do not dissolve upon heat-up.

### Lubricant Specifications and Requirements

Gas turbine lubricant requirements are set by turbine manufacturers such as those shown in Table 1. These specifications delineate sets of required laboratory tests, each measuring a relevant lubricant property. The tests and limits set in these specifications are meant to preclude lubricant related failures from impeding turbine operation. They vary per the individual design and running conditions of each model of gas turbine and the level of performance needed. The specification limits are frequently reviewed and are upgraded as necessitated by the recent changes in turbine design.

Several lubricant industry organizations have also assembled gas turbine lubricant specifications, for example:

- ASTM D 4304
- DIN 51515
- British Standard 489
- ISO 8068
- JIS K 2213
- China National Standard GB11120-89

## Gas Turbine Lubricant Tests and Properties

### Oxidative Stability

Due to the extreme high temperature exposure, thermal and oxidative stability are the most critical properties of gas turbine oils. While bulk oil temperatures are usually low, there can be localized hot spots such as in bearings, gas seals, and throttle control mechanisms that can cause accelerated thermal/oxidative degradation. The hostile environment of a turbine provides ideal ingredients for lubricant oxidation – heat, aeration, water, and metal catalysis. Hydrocarbon oils oxidize via a free radical chain reaction (Rasberger 1997). Among the products of oxidative degradation are acidic materials plus polymerized products such as sludge and varnish deposits (Fitch and Gebarin 2008). Even small amounts of sludge and varnish can lead to problems and malfunctioning of critical turbine systems, leading to unscheduled shut-downs. Determination of a gas turbine lubricant’s potential oxidative stability is thus essential. A variety of accelerated laboratory oxidation tests, such as those shown in Table 2, have been utilized to simulate the oxidative exposure of gas turbine oils. Pass limits for these tests have generally been increased to insure good operation in today’s more demanding running conditions. Demonstrated performance in a combination of several of these tests is usually required to build sufficient confidence in a lubricant’s stability.

*Turbine Oil Stability Test (TOST)* – Estimates the oxidation stability of lubricants in the presence of water, oxygen, copper, and iron accelerants at 95°C in a glass tube. The time required to reach an acid number of 2.0 is defined as the *oxidation lifetime*. This test, originally adopted in 1947, remains one of the most important performance criteria for gas turbine oils since it correlates well with demonstrated turbine oil field service life. Several variations of the technique have been added over the years, all utilizing the original test cell. Early turbine oils had TOST lifetimes of 1–2,000 h. With recent

**Gas Turbine Engine Lubricants, Table 2** Gas turbine oxidation stability tests

Test	Designation(s)	Typical specification requirements
Turbine oil stability test (TOST)	ASTM D 943, DIN 51587, ISO 4263, IP 157, JIS K 2514	3,000–7,000 h (min)
TOST sludge	ASTM D 4310	100–200 mg (max)
Rotary pressure vessel oxidation test (RPVOT)	ASTM D 2272, IP 229, JIS K 2514	300–1,200 min (min)
“Modified” RPVOT	GEK-32568A	80–85% of original value (min)
Oxidation corrosion stability (OCS)	ASTM D 4636 Proc. 2, FTMS 791c 5308.6, DIN 51394	
Dry TOST life	<sup>a</sup>	500–1,000 h (min)
Sludge		100 mg/kg sludge when 50% or 25% of original RPVOT is reached

<sup>a</sup>Currently MHI procedure; ASTM is creating a standard test method

improvements in base oil quality (e.g., Group II base stocks) and additive quality, it is not unusual to reach 10,000 h – the limit of the test. Many oils are marketed with even higher TOST values claimed, but the test needs to be modified to run to such durations.

**TOST sludge** – Lubricant is exposed to water, oxygen, copper, and iron at 95°C in a glass tube for 1,000 h; the sample is filtered and the amount of insolubles formed is determined by filtration. The total amount of copper corrosion found in the oil, water, and sludge phases may also be determined.

**Rotary Pressure Vessel Oxidation Test (RPVOT)** – Estimates the oxidation stability of lubricants in the presence of water and a copper coil at 150°C in a rotating oxygen pressurized vessel. The number of minutes required to achieve a 25 psi pressure drop is the oxidation stability of the fluid. This test is best used for batch acceptance of production lots or to assist in tracking the oxidation life of in-service oils. The key benefit of the RPVOT test is its relative speed. However, this is accomplished by introducing an unrealistic feature not present in turbine service – i.e., high pressure (plus water at 150°C). Despite the fact that the RPVOT is not intended to be a substitute for the TOST (often the two do not correlate) or to be used for comparing the estimated

service lives of new oils of different compositions, it has made its way into most gas turbine oil specifications. Retention of RPVOT value during use is considerably more important than the actual initial value. As with the TOST, RPVOT values have increased significantly with use of Group II oils; values of 2,000–3,000 min can easily be achieved with formulations utilizing these base oils.

**Modified RPVOT** (better designated as RPVOT Retention) – Gas turbines operate at temperatures high enough to volatilize low-molecular weight additives that demonstrate high RPVOT values (i.e., DBPC – 2,6-ditertiary butyl paracresol). In actual use, these additives have poor durability in a gas turbine. The modified RPVOT technique was thus introduced by GE to correct for this deficiency. The RPVOT is first determined on the fresh lubricant. Another portion of the lubricant is purged with nitrogen for 48 h at 121°C. At the conclusion of the treatment, the RPVOT is rerun and compared with the original value. Minimum retention values, expressed as a percentage of the original, are required by specifications.

**Oxidation and Corrosion Stability** – This procedure stresses the lubricant in the presence of five metal catalysts (steel, Cu, Al, Ag, and Mg). Air is blown through 175°C oil for 72 h, after which sludge formation, acid number, and viscosity increase and metal corrosion is determined.

**Dry TOST** – It was noted (Chasan et al. 2008) that gas turbine formulations can easily last 10,000 h in TOST with low acid numbers, but are heavily loaded with sludge at that point (despite having very low sludge values in the standard 1,000-h TOST sludge test). Mitsubishi Heavy Industries (MHI) also noted this sludge formation in gas turbine oils in field use (Yano et al. 2004). The problem is, in part, due to the increased use of new Group II base oils. These fluids are inherently more oxidatively stable than Group I base oils, but with their low aromatic content, they, and turbine oils formulated with them, have a significantly reduced ability to solubilize high-molecular-weight, polar oxidation products. MHI therefore created a new variation of the TOST test to screen new formulations for protection against this behavior. The traditional TOST apparatus and metal catalysts are used, but water is excluded and the test is run at 120°C. MHI gas turbine lubricant specifications stipulate minimum lifetimes and limit the amount of sludge that may be present when the oil reaches certain levels of degradation (as measured by percentage of initial RPVOT value). The test is difficult to pass and has required improved oxidative protection for gas turbine oils.



### Miscellaneous Other Oxidation Tests

**Varnish Precursor Tests** – Increased use of highly refined, low-solvency Group II base oils has also led to increased instances of varnish formation in gas turbines. In addition to standard oxidation mechanisms, occurrences of electrostatic discharges linked to oil filtration have been associated with varnish formation in gas turbines (Sasaki et al. 1999a, b). The increased presence of varnish can lead to filter plugging, bearing wear, reduced heat exchanger performance, and poor servo-valve response. Several types of tests have been proposed to screen for oil degradation products that present themselves as varnish precursors. Most involve filtration of the turbine oil through a filter, solvent wash of the filter, and spectrometric rating of the filter patch. One such test is the membrane patch colorimetry test.

**High Pressure Differential Scanning Calorimetry (HPDSC)** – ASTM D 6186 is an increasingly used rapid thin film isothermal procedure used to determine the oxidation induction time of fresh and used turbine oils.

**CIGRE (IP 280)** – Required only by BS 489. Oxygen is passed through a heated oil sample containing soluble metal (iron and copper) oxidation catalysts. Total oxidation products (sludge plus volatile acid products plus oil acidity) are determined.

A variety of other turbine oil properties are also listed in manufacturers' specifications and are defined by the test methods and limits specified. These are summarized in Table 3.

### Corrosion Protection

Protection of ferrous parts from corrosion caused by moisture and acidic degradation products is required by all gas turbine oil specifications. Dissolved metal ions can also catalyze oil oxidation, so their generation must be controlled. In the ASTM D 665 corrosion test, a polished steel rod is immersed in a stirred beaker with oil and (distilled or sea) water at 140°F, usually for 4 or 24 h. At the conclusion of the test the rod is visually rated for the presence of corrosion.

Protection is also needed to preserve brass-, bronze-, and copper-containing parts of the turbine. The ASTM D 130 copper corrosion test measures corrosion of a copper specimen immersed in 100°C oil for a period of three hours. The specimen is then visually rated for excessive corrosion.

### Antiwear and Load Carrying

The FZG gear rig test, ASTM D 5182, is the procedure used to insure that adequate "extreme pressure (EP)" load carrying protection is present for lubricants used in

Gas Turbine Engine Lubricants, Table 3 Other gas turbine oil properties

Property	Test designation(s)	Typical specification requirements
Ferrous corrosion	ASTM D665(A and B), [DIN 51585 (expired)], ISO 7120, IP 135	Pass
Copper corrosion	ASTM D130, DIN 51759 Part 1, ISO 2160, IP 154	1–2 maximum
Antiwear/extreme pressure (load carrying)	ASTM D 5182, DIN 51354	8–9 failure load stage minimum
Foam resistance	ASTM D892, DIN 51566/53566, ISO 6247, IP 146	20–300 ml maximum
Air release	ASTM D 3427, IP 313	4–5 min maximum
Demulsibility	ASTM D1401, DIN 51599, ISO 6614	15–30 min
Viscosity	ASTM D445, DIN 51366/51561/51562/51569, ISO 3104, IP 71	ISO 32, ISO 46, ISO 68, ISO 100
Viscosity index	ASTM D 2270, ISO 2909, IP 226	90–95 minimum
Acid/Neutralization number	ASTM D 974, DIN 51558 Part 1, ISO 6618, IP 139/ D 664, IP 177	0.20 (0.30 for EP oils) maximum
Pour point	ASTM D 97, DIN 51597, ISO 3016, IP 15	+10 to –12°C maximum
Flash point	ASTM D 92, ISO 2592, IP 36	160–215°C minimum

certain geared gas turbines. These EP turbine oils generally must exceed a failure loading stage of 8 or 9 to be eligible for service in such turbines.

### Foam Resistance

Poor base oil quality, additive incompatibility, degradation products, and contaminants can cause excessive foam that will compromise the performance of a gas turbine lubricant. Among the issues that may arise from excessive foam are increased oxidation, additive depletion, reduced heat transfer coefficient, overflow of the lubricant, microdieseling, and increased wear caused by reduced fluid viscosity. The ASTM D 892 foam measurement test is conducted in three sequences during which air is blown through a diffusion stone into the lubricant. After 5 min, the air flow is terminated and the volume of foam present

at that point and after 10 min of settling is measured. Specification requirements vary from 20 to 300 ml foam allowed.

### Air Release

Air release is a measure of the ability of a lubricant to dispel entrained air by releasing air bubbles. The same factors that lead to increased foaming tendencies can also cause air to be entrained in the lubricant. Entrained air can cause accelerated oxidation, reduced heat transfer, cavitation, and sponginess, resulting in loss of hydraulic response. Compressed air is blown through the sample for 7 min, after which, the time required for air entrained in the oil to decrease to 0.2% is measured as the separation time. Specifications for gas turbine oils usually have a 4- or 5-minute maximum air release requirement.

### Demulsibility

A turbine oil must have the ability to separate from water. Entrained water can accelerate oil degradation, chemical corrosion, bearing fatigue, and additive extraction. Certain additives and contaminants can adversely affect the demulsibility properties. In the standard ASTM D 1401 test, oil is vigorously stirred for 5 min with an equal quantity of water. Upon conclusion of the stirring, the time and nature of the separation of the phases is measured.

While exposure to moisture is normally a major concern for steam turbine lubricants, gas turbine oils will also be exposed to moisture condensation during cyclic operation.

### Viscosity

Manufacturers will usually specify either an ISO 32, ISO 46, ISO 68, or ISO 100 viscosity grade turbine oil as appropriate for a specific turbine. The recommendation will depend on shaft speed, bearing clearance, oil circulation rate, gear tooth load, gear and bearing temperatures, and oil reservoir temperature (Okasaki et al. 2005). It is critical for the appropriate viscosity fluid to be used to create a hydrodynamic oil film under the turbine's operating conditions.

Viscosity index (VI) is a measure of the turbine oil's ability to retain its viscosity when heated and is also detailed in manufacturers' specifications. The less fluid thinning that occurs at higher temperature, the higher the viscosity index (VI) of the lubricant will be.

### Acid/Neutralization Number

The presence of acidic constituents in a gas turbine lubricant (from base oil and/or additives) is determined by

a titration with alcoholic KOH. In order to minimize acid-induced corrosion and oxidation and interactions within the turbine oil formulations, the neutralization number or total acid number of the fresh lubricant is usually limited to 0.20 maximum and to 0.3 maximum for some EP gas turbine oils. Either ASTM D 664 or D 974 titration is required by most specifications.

### Pour Point and Flash Point

To insure safe and continued operation of the gas turbine, pour point and flash point requirements are set according to the temperature exposures anticipated for the lubricant in the specific turbine.

### Gas Turbine Lubricant Formulations

Gas turbine lubricants are composed of two components: base oil and additives. To achieve current required performance levels, highly refined base oils must be used in combination with a carefully balanced, thermally stable additive package.

Lubricant base oils have been categorized by the American Petroleum Institute (API) into the five groups shown in Table 4, according to sulfur and saturates content and viscosity index.

Earlier gas turbine oils were made from Group I base oils. Simple rust and oxidation (R&O) inhibiting lubricants were adequate to meet the requirements. Fortunately, the evolving requirements for increased turbine oil performance have been accompanied by the emergence of new generations of base fluids and additive technologies (Swift et al. 2001). Group II and Group III base oils are now produced with a significant reduction in highly reactive aromatic content, as well as detrimental sulfur and basic nitrogen compounds. The result is an improvement in inherent oxidative stability of lubricant formulations using these oils. Additionally, additive response, color, and air release properties are also enhanced. On the other hand, the removal of sulfur from these oils has eliminated

Gas Turbine Engine Lubricants, Table 4 Base oil types

Group	Sulfur content, wt%		Saturate content	VI
I	>0.03	And/or	<90	80–119
II	≤0.03	And	≥90	80–119
III	≤0.03	And	≥90	≥120
IV	All polyalphaolefins (PAOs)			
V	All stocks not included in groups I–IV (pale oils and non-PAO synthetics)			

**Gas Turbine Engine Lubricants, Table 5** Gas turbine lubricant additives

Additive type	Typical additives	Typical treat levels
Antioxidant	Alkylated amines, hindered phenols, organic sulfur compounds, phosphites – and combinations thereof	0.20–1.0%
Rust/corrosion inhibitor	Alkylated succinic acid esters, half esters and amines, amine phosphates	0.03–0.2%
Demulsifier	Block copolymers of ethylene and propylene oxides	0.01–0.10%
Copper passivator/metal deactivator	Triazoles, thiadiazoles	0.02–0.2%
Antifoamant	Polysiloxanes, polyacrylates, polymethacrylates	5–100 ppm
Pour point depressant (occasional)	Polymethacrylates	0.1–0.5%
Extreme pressure/load carrying (for EP oils)	Sulfurized olefins, fatty acid esters, phosphorothionates, dithiocarbamates, amine phosphates	0.03–0.5%

some naturally occurring antioxidants. Additionally, as noted above, the ability of such highly paraffinic oils to solubilize sludge and varnish is severely compromised. Thus, when using Group II or Group III base oils to formulate gas turbine oils, different additives as well as increased additive treat levels may be required. Formulation strategies must be less focused on acid number and viscosity control and more on controlling generation of sludge and varnish. Additionally, since they are poorer solvents than Group I oils, limited additive solubility in these oils must be considered.

Suitable additives are selected to enable the gas turbine lubricant to meet the fluid requirements delineated by the specifications in Table 1 as measured by the procedures listed in Tables 2 and 3. A typical gas turbine lubricant will contain a total of 0.5–1.5 wt% of the types of additives found in Table 5. A common formulation strategy is to capitalize on the use of synergistic combinations. This will maximize performance and minimize both treat levels (cost) and risk of solubility issues. The classic example used in formulating gas turbine oils is the combination of

aminic and phenolic radical scavenging antioxidants. The phenolic antioxidant will help regenerate the aminic antioxidant and, together, they efficiently disrupt the oxidation chain reaction. The result is extended lubricant lifetime. Additionally, primary, radical scavenging antioxidants can be mixed with secondary (sulfur and phosphorus containing) peroxide decomposing antioxidants to boost performance. Optimal antioxidant ratios will vary with selection of base oil.

### In-Service Turbine Lubricant Monitoring

“Standard Practice for In-Service Monitoring of Mineral Turbine Oils for Steam and Gas Turbines,” ASTM D 4378, was created in 1984 with joint input from turbine equipment manufacturers, oil suppliers, and power-generation industry representatives. It is a very effective guide for proper monitoring of in-service turbine lubricants to ensure long-term, trouble-free operation. Lubricant sampling and testing schedules are delineated and warning limits, assistance with data interpretation, and recommended action steps are included. Oils will not remain unchanged during use in a turbine. Additives will often be transformed into other active intermediate species before their useful life is exhausted. Thermal and oxidative deterioration will eventually occur as well as additive depletion and contamination. Monitoring trends of changes in the used lubricant’s properties, as compared with the fresh oil, will assist in identifying signs of lubricant deterioration before damage can occur. Mild degradation can be handled with moderate remedial actions such as increased monitoring, readditization or top-up; extreme deterioration will require a complete oil drain and change-out.

Depletion of a turbine oil’s additives may occur because the additive is consumed while performing its function, or because it is removed by volatilization, migration into water, adsorption onto (wear) debris, or reaction with contaminants.

The following evaluations are recommended by ASTM D 4378 for monitoring in-service gas turbine lubricants’ condition:

- Visual inspection (foam, clarity, particles, odor)
- Viscosity
- Acid number
- Appearance
- Color
- Cleanliness (particle count, wear metals)
- Oxidation reserve (either by RPVOT, RULER [voltammetry], or FTIR). It has been noted that RPVOT may be an inadequate indicator of the

presence of sludge and varnish degradation products (Livingstone et al. 2008 and Sasaki et al. 2008).

Most of these inspections will monitor the onset of lubricant deterioration. Viscosity check also insures that there has been no contamination (e.g., via makeup) with a lubricant of improper viscosity grade. The cleanliness inspection will check for the ingress of dust, ash, or other foreign objects in the oil as well as wear particles. A test for varnish precursors may also be beneficial. Flash point, foam, and rust may also be evaluated in certain circumstances. Data from a combination of several of these monitoring tests should be utilized for determining what actions might be suitable. When make-up oil is used that is different from the original formulation, compatibility testing with the original oil must be conducted before use.

The eventual turbine lubricant lifetime will be determined by a combination of the following factors:

- Type and design of system
- Condition of system on start-up
- Original oil quality
- System operating conditions
- Contamination
- Oil makeup rate
- Handling and storage

The amount and frequency of a turbine's makeup oil will also play a significant role in determining the lifetime of the oil charge. Makeup can vary from 5% to 30% depending on the turbine running conditions. A high makeup rate could extend the service life of the oil. When top-up is conducted with a gas turbine lubricant other than the original charge, it is necessary to run functional and compatibility tests to insure the health of the resulting lubricant mix. These can include a performance check such as RPVOT and visual test for haze or precipitate formation.

With proper monitoring, a steam turbine oil can last from 10 to 20 years. Gas turbine lubricant lifetime, on the other hand, will vary significantly with the severity of the turbine's operation. A low-severity gas turbine could exceed 100,000 h lubricant service, while high-severity units will attain only 10,000–30,000 h operation with their lubricant (Robertson 1989).

### Fire-Resistant Turbine Lubricants

To reduce the risk of oil fires in steam and gas turbines, aryl phosphate ester fluids have been used as alternates to mineral oils. Of 175 turbine fires reported between 1930 and 1983, 121 involved oil as a primary or secondary source. The costs of turbine fires can be very high, outage

costs are considerable, and safety risks exist. Triarylphosphate esters have autoignition temperatures in the region of 550–590°C (~200° higher than mineral oils) and possess inherent self-extinguishing properties. There are, though, some drawbacks to the use of phosphate esters, e.g., much lower viscosity indices, higher densities (requiring more powerful pumps), hydrolysis of the phosphate, and incompatibility with conventional paints and seals. This requires that systems be specially designed for use with phosphates since modifications may be needed to accommodate these properties. There is sufficient turbine history with phosphate lubricants to demonstrate their viability (Phillips 2001). Higher fluid costs and the general conservativeness of the power generation industry have limited the utilization of these fluids to date. The continuing trend to higher operating temperatures to boost turbine efficiency may see an increasing reliance on this type of fluid.

### Aviation Gas Turbine Lubricants

Military and commercial jet aircraft may use a gas turbine to drive a rotor (helicopter) or propeller (turboprop) or to create turbothrust (ramjet). Aviation gas turbines require the use of synthetic base fluids to meet the thermal and oxidative operational demands placed on their lubricants. Carboxylate esters from trimethylolpropane or pentaerythritol reacted with short-chain fatty acids have traditionally been used as base fluids for aviation gas turbine oils. By varying the constituent acid and alcohol, esters can be tailor-made to suit specific applications. They provide stable lubrication over a much wider temperature regime than can be obtained with mineral oil-based fluids (e.g., –50°C to 204°C). Low volatility, corrosion protection, and load-carrying capabilities are also required. The US military specifications governing the requirements of these fluids are:

MIL-L-7808  
MIL-PRF-23699  
MIL-PRF-7808  
MIL-L-27502

A key oxidation and corrosion stability test for these fluids is ASTM D 4636/FTMS 5307/FTMS 5308, often run at hotter temperatures than for mineral oil-based formulations. Trends to hotter, more fuel-efficient engines have placed an added emphasis on also measuring coking tendencies of these lubricants (Gschwender et al. 2001).

Compact, lighter aero-derived gas turbines are commonly used for both land and sea operations. Essentially they are aviation-type gas turbines that have been scaled down in size for land-based use. Applications include

power generation in remote areas (e.g., offshore platform rigs), petro-chemical pipeline operations, and marine propulsion. These are usually high temperature units and their severe lubrication requirements are also satisfied by synthetic aviation gas turbine oils.

## References

- D. Chasan, S. DiFrancesco, M. Ribeaud, Oxidation testing of long-life turbine oils fluids. Can we do better? *Oxid. Test. Turbine Oils*, ASTM STP **1489**, 95–102 (2008)
- J. Fitch, S. Gebarin, Review of degradation mechanisms leading to sludge and varnish in modern turbine oil formulations. *ASTM STP* **1489**, 54–63 (2008)
- L. Gschwender et al., Advanced high temperature air force turbine engine oil program. *ASTM STP* **1407**, 17–24 (2001)
- J.B. Hannon, How to select and service turbine oils. *Mach. Lubric.* **7**, 36–42 (2001)
- G. Livingstone, B. Thompson, M. Okazaki, Physical, performance and chemical changes in turbine oils from oxidation. *ASTM STP* **1489**, 27–44 (2008)
- M. Okasaki, R. Mesquita, P.J. Sweeny, T. Opstal, Power Generation Lubricants: Gas and Steam Turbines, Lubrication (Chevron Products Company, San Ramon, CA (2005), pp. 1–12
- D. Phillips, The use of fire-resistant turbine lubricant: Europe looks to the future. *ASTM STP* **1407**, 1–16 (2001)
- M. Rasberger, Oxidative Degradation and Stabilization of Mineral Oil Based Lubricants, in *Chemistry and Technology of Lubricants*, ed. by R. Motier, S. Orszulik (Blackie Academic and Professional, London, 1997), pp. 98–143
- R.S. Robertson, Background and development of ASTM D 4378: practice for in-service monitoring of mineral turbine oils for steam and gas turbines. *ASTM STP* **1021**, 3–24 (1989)
- A. Sasaki, S. Uchiyama, T. Yamamoto, Generation of static electricity during filtration. *Lubric. Eng.* **55**, 14–21 (1999a)
- A. Sasaki, S. Uchiyama, T. Yamamoto, Free radicals and oil autooxidation due to spark discharges of static electricity. *Lubric. Eng.* **55**, 24–27 (1999b)
- A. Sasaki, S. Uchiyama, M. Kawasaki, Varnish formation in gas turbine oil systems. *ASTM STP* **1489**, 103–114 (2008)
- A.N. Smith, Heavy Duty Gas Turbines, in *Handbook of Lubrication*, ed. by E.R. Booser, vol. I (CRC Press, Boca Raton, 1987), pp. 107–119
- S. Swift, K.D. Butler, W. Dewald, Turbine oil quality and field application requirements. *ASTM STP* **1407**, 39–52 (2001)
- A. Yano et al., Study on sludge formation during the oxidation process of turbine oils. *Tribol. Trans.* **47**, 111–122 (2004)

## Gaseous Cavitation

- [Cavitation Formation and Modeling](#)

## Gaseous Squeeze Film

- [Squeeze Film Gas Lubrication](#)

## Gas-Lubricated Face Seal

- [Dry Gas Seals](#)

## Gauge Corner Cracking (GCC)

- [Rolling Contact Fatigue \(RCF\)](#)

## Gear Bending Fatigue Failure and Bending Life Analysis

ROBERT ERRICHELLO

GEARTECH, Townsend, MT, USA

### Synonyms

[Bending fatigue](#); [Bending fatigue resistance](#); [Bending strength](#); [Root stress](#)

### Definition

Bending fatigue is caused by repeated bending stresses that exceed the local fatigue strength in the tensile root fillet of a gear tooth. A fatigue crack initiates at the surface of the root fillet and propagates into the gear tooth normal to the root fillet surface. Near the centerline of the tooth, the crack turns and grows toward the compression root fillet. The crack trajectory leaves a convex surface on the broken tooth fragment and a concave surface on the gear body. ANSI/AGMA 1010-E95 (1995) describes the characteristics of the failure mode. [Figure 1](#) shows bending fatigue failure on a spur gear tooth.

### Scientific Fundamentals

Gear tooth load capacity based on bending fatigue resistance is calculated according to industrial standards such as ANSI/AGMA 2101-D04 (2004) and ISO 6336-3:1999 (1999). Although bending stress is a major factor governing bending fatigue, the gear material exerts an equally important influence. Therefore, to prevent bending fatigue, the metallurgical properties of the gear should conform to the requirements of industrial standards such as ANSI/AGMA 2101-D04 (2004).

### Bending Stress Analysis

Industrial standards (ANSI/AGMA 2101-D04 2004; ISO 6336-3:1999 1999) calculate bending stress with an





**Gear Bending Fatigue Failure and Bending Life Analysis,**  
**Fig. 1** Bending fatigue failure on a spur pinion

equation similar to (1). See references (AGMA 925-A03 2003; ANSI/AGMA 2101-D04 2004) for information on the variables used in (1) and (2).

$$\sigma_F = \frac{w_n}{bG} \quad (1)$$

$$w_n = \frac{F_{wn} K_O K_H K_V}{L_{\min}} \quad (2)$$

where

$\sigma_F$	Bending stress, N/mm <sup>2</sup>
$w_n$	Normal unit load, N/mm
$b$	Face width, mm
$G$	Geometry factor accounts for shape of gear tooth profile and root fillet, location of load on the tooth profile that causes maximum bending stress, and stress concentration in the root fillet, dimensionless
$F_{wn}$	Total load normal to the gear tooth profiles at the point that causes maximum bending stress, N
$K_O$	Overload factor accounts for externally applied loads induced by dynamics of the gear system, dimensionless
$K_H$	Load distribution factor accounts for non-uniform distribution of load along the lines of contact between mating gear teeth, dimensionless
$K_V$	Dynamic factor accounts for internally generated gear tooth loads induced by non-uniform meshing action of gear teeth, dimensionless
$L_{\min}$	Minimum contact length, mm

### Bending Life Analysis

Industrial standards (ANSI/AGMA 2101-D04 2004; ISO 6336-3:1999 1999) use empirical methods to relate

calculated values of bending stress to empirical values of allowable bending stress with an equation similar to (3). See reference (ANSI/AGMA 2101-D04 2004) for information on the variables used in (3).

$$S_F = \frac{\sigma_{FP} Y_N}{\sigma_F Y_\theta Y_Z} \quad (3)$$

where

$S_F$	Safety factor accounts for uncertainties or statistical variations in design analysis, material properties, and manufacturing tolerances, dimensionless
$\sigma_{FP}$	Allowable bending stress for a baseline design life such as $10^7$ cycles, N/mm <sup>2</sup>
$Y_N$	Stress-cycle life factor adjusts the allowable stress for a design life different than the baseline design life, dimensionless
$\sigma_F$	Bending stress, N/mm <sup>2</sup>
$Y_\theta$	Temperature factor accounts for reduction of fatigue strength at high temperature, dimensionless
$Y_Z$	Reliability factor accounts for statistical variation in fatigue strength, dimensionless

### Key Applications

Industrial standards such as (ANSI/AGMA 2101-D04 2004; ISO 6336-3:1999 1999) provide fundamental rating formulas to evaluate gear tooth load capacity. Application standards tailor the fundamental formulas to particular applications such as wind turbine gears (ANSI/AGMA/AWEA 6006-A03 2003) and marine gears (ANSI/AGMA 6032-A94 1994).

### Cross-References

- [Fatigue](#)
- [Gear Materials](#)
- [Shot Peening](#)

### References

- AGMA 925-A03, *AGMA Information Sheet – Effect of Lubrication on Gear Surface Distress* (AGMA, Alexandria, 2003)
- ANSI/AGMA 1010-E95, *AGMA Standard – Appearance of Gear Teeth – Terminology of Wear and Failure* (AGMA, Alexandria, 1995)
- ANSI/AGMA 2101-D04, *AGMA Standard – Fundamental Rating Factors and Calculation Methods for Involute Spur and Helical Gear Teeth* (AGMA, Alexandria, 2004)
- ANSI/AGMA 6032-A94, *AGMA Standard- Standard for Marine Gear Units: Rating* (AGMA, Alexandria, 1994)
- ANSI/AGMA/AWEA 6006-A03, *AGMA Standard – Standard for Design and Specification of Gearboxes for Wind Turbines* (AGMA, Alexandria, 2003)
- ISO 6336-3:1999, *ISO International Standard – Calculation of Load Capacity of Spur and Helical Gears – Part 3: Calculation of Tooth Bending Strength* (ISO, Geneva, 1999)

## Gear Contact Temperature and Scuffing Risk Analysis

ROBERT ERRICHELLO

GEARTECH, Townsend, MT, USA

### Synonyms

Galling; Seizure; Severe adhesion

### Definition

Scuffing is localized damage caused by solid-phase welding between surfaces in relative motion. It is accompanied by transfer of metal from one surface to another due to welding and subsequent tearing. ANSI/AGMA 1010-E95 (1995) describes the characteristics of the failure mode. AGMA 925-A03 (2003) gives information on effects of lubrication on gear surface distress. Figure 1 shows scuffing on a spur pinion.

### Scientific Fundamentals

Gear tooth load capacity based on the probability of scuffing is calculated according to industrial standards such as AGMA 925-A03 (2003), which calculates scuffing risk using Blok's critical temperature theory (Blok 1937). Although contact temperature is a major factor governing scuffing, the properties of the lubricant and its additives exert a strong influence (Errichello 1992). Therefore, to help prevent scuffing, the physical and chemical properties of the lubricant should conform to the requirements of ANSI/AGMA 9005-D94 (1994). It is especially important

that the lubricant have adequate antiscuff additives to ensure adequate scuffing resistance.

### Contact Temperature Analysis

AGMA 925-A03 (2003) calculates contact temperature with equation (1). See reference (AGMA 925-A03 2003) for information on the variables used in (1) through (6).

$$\theta_B = \theta_M + \theta_{fl} \quad (1)$$

$$\theta_{fl} = \frac{31.62 K \mu_m X_\Gamma w_n |v_{r1} - v_{r2}|}{(b_H)^{1/2} \left( B_{M1} (v_{r1})^{1/2} + B_{M2} (v_{r2})^{1/2} \right)} \quad (2)$$

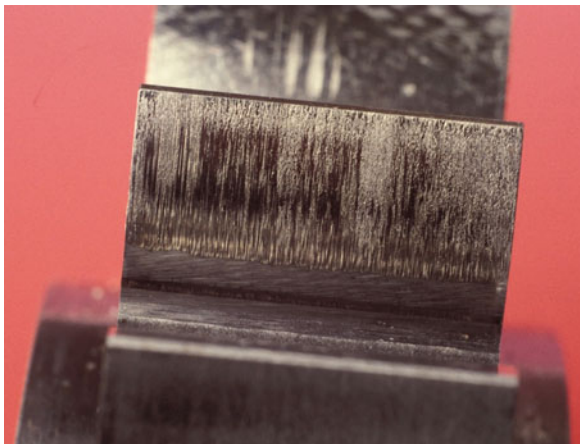
$$w_n = \frac{F_{wn} K_O K_H K_V}{L_{min}} \quad (3)$$

$$b_H = \left( \frac{8 X_\Gamma w_n \rho_n}{\pi E_r} \right)^{1/2} \quad (4)$$

$$E_r = 2 \left( \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2} \right)^{-1} \quad (5)$$

where

$\theta_B$	Contact temperature, °C
$\theta_M$	Tooth temperature, °C
$\theta_{fl}$	Flash temperature, °C
$K$	Flash temperature constant accounts for distribution of frictional heat over the contact band, $2b_H$ , dimensionless
$\mu_m$	Mean coefficient of friction, dimensionless
$X_\Gamma$	Load sharing factor, dimensionless
$w_n$	Normal unit load, N/mm
$v_{r1}, v_{r2}$	Rolling velocity (pinion, gear), m/s
$b_H$	Semi-width of the Hertzian contact band, mm
$B_{M1}, B_{M2}$	Thermal contact coefficient (pinion, gear), $N/[mm \ s^{0.5} \ K]$
$F_{wn}$	Total load normal to the gear tooth profiles at the operating pitch point, N
$K_O$	Overload factor accounts for externally applied loads induced by dynamics of the gear system, dimensionless
$K_H$	Load distribution factor accounts for non-uniform distribution of load along the lines of contact between mating gear teeth, dimensionless
$K_V$	Dynamic factor accounts for internally generated gear tooth loads induced by non-uniform meshing action of gear teeth, dimensionless
$L_{min}$	Minimum contact length, mm
$\rho_n$	Normal relative radius of curvature, mm



Gear Contact Temperature and Scuffing Risk Analysis,  
Fig. 1 Scuffing on a spur pinion

$E_r$	Reduced modulus of elasticity, N/mm <sup>2</sup>
$\nu_1, \nu_2$	Poisson's ratio (pinion, gear), dimensionless
$E_1, E_2$	Modulus of elasticity (pinion, gear), N/mm <sup>2</sup>

Scuffing Risk Analysis

AGMA 925-A03 (2003) uses empirical methods to relate calculated values of maximum contact temperature to empirical values of scuffing temperature with equation (6), which uses a Gaussian distribution to calculate the probability of scuffing. See reference (AGMA 925-A03 2003) for information on the variables used in (6).

$$x = \frac{\theta_{Bmax} - \theta_s}{\sigma_{\theta s}} \tag{6}$$

where

$x$	Standard normal variable, dimensionless
$\theta_{Bmax}$	Maximum contact temperature, °C
$\theta_s$	Mean scuffing temperature, °C
$\sigma_{\theta s}$	Standard deviation of mean scuffing temperature, °C

Key Applications

Industrial standards such as (AGMA 925-A03 2003) provide fundamental rating formulas to evaluate gear tooth load capacity. Application standards tailor the fundamental formulas to particular applications such as wind turbine gears (ANSI/AGMA/AWEA 6006-A03 2003) or marine gears (ANSI/AGMA 6032-A94 1994).

Cross-References

- Gear Lubricants
- Gear Lubrication

References

AGMA 925-A03, *AGMA Information Sheet – Effect of Lubrication on Gear Surface Distress* (AGMA, Alexandria, 2003)

ANSI/AGMA 1010-E95, *AGMA Standard – Appearance of Gear Teeth – Terminology of Wear and Failure* (AGMA, Alexandria, 1995)

ANSI/AGMA 6032-A94, *AGMA Standard – Standard for Marine Gear Units: Rating* (AGMA, Alexandria, 1994)

ANSI/AGMA 9005-D94, *Industrial Gear Lubrication* (AGMA, Alexandria, 1994)

ANSI/AGMA/AWEA 6006-A03, *AGMA Standard – Standard for Design and Specification of Gearboxes for Wind Turbines* (AGMA, Alexandria, 2003)

H. Blok, *Les Temperature de Surface dans les Conditions de Graissage sans Pression Extreme*, Second World Petroleum Congress, Paris, June 1937

R. Errichello, Friction, lubrication, and wear of gears, in *ASM Handbook*, vol. 18 (ASM International, Materials Park, 1992), pp. 535–545

Gear Cutting

- Gear Manufacturing

Gear Cutting Tools

CHARLES D. SCHULTZ  
Beyta Gear Service, Winfield, IL, USA

Synonyms

Broaches; Carbide insert cutters; Coated cutters; Form cutters; Hobs; Insert cutters; Rack shaper cutters; Shaper cutters

Definition

Gear cutting tools are the implements used to form the tooth spaces in a gear.

Scientific Fundamentals

Early gears were made from wood and the tooth spaces were carved with the same stone, bone, or soft metal implements used to make or decorate other wooden artifacts. As gears transitioned to metal, more specialized tools were needed, especially as the gear blanks became harder. Improvements in gear machinery and processing closely parallel improvements in the ability to make harder and more durable tools.

The tooth spaces in early clock gears were hand filed rather than cut. Surviving implements include triangular-shaped files and indexing plates, a methodology that required a high degree of operator skill to make functional gears. This was impractical on larger teeth, and for many years larger metal gears were restricted to cast teeth that could be hand filed if needed to obtain smoother power transmission.

By 1672, Englishman Robert Hooke was using a hand-powered cutting machine to form gear teeth. His cutter was basically a file-like surface wrapped around a wheel with a cross section that matched the space to be cut. Once a space was cut, the blank was indexed to the next position and the process was repeated. As better tool steels became available, that “rotary file” evolved into the form milling cutter still in use today. As early as 1825, machine power was being used in Switzerland to cut tooth spaces. More power was required better tool steel but allowed for faster cutting. It is a dynamic still at work today. Index accuracy

gradually improved over the years as the milling mechanism was connected via a gear train to the index mechanism to produce the first dedicated gear-cutting machines.

Gear accuracy is limited by the precision of the cutter profile when the tooth space is form milled. To get better gears, it was recognized that the tooth profiles needed to be “generated” by controlled motion of the cutter in time with the part. This generating action reduced the need to make a precise curve on the cutting tool. Englishman Joseph Saxton was using a straight-sided tool to generate involute curved tooth flanks by 1842. This property of the involute curve has contributed to its ongoing popularity as a tooth form.

Once “generating” became understood, innovators turned their attention to making gears better, faster, quieter, and longer lasting.

## Key Applications

### Form Cutters

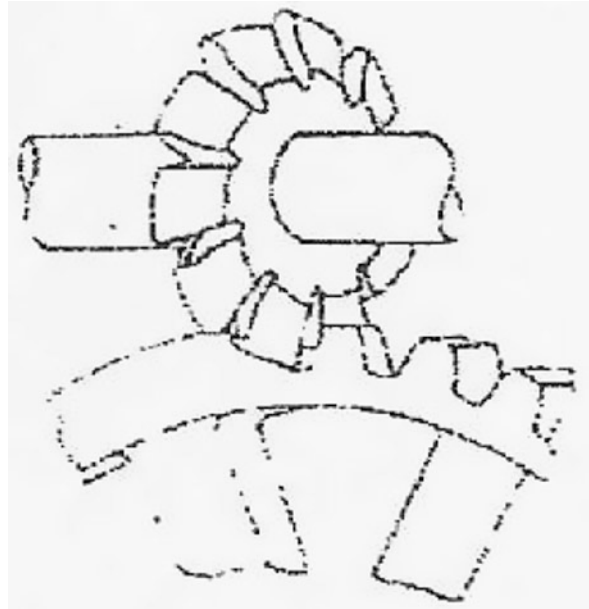
The term “form cutter” refers to the cross section of the cutter matching the “form” or shape of the desired tooth space (Fig. 1). With the proper cutter and a suitable indexing attachment, a standard milling machine can produce spur gears of commercial quality [AGMA Q6]. As it was expensive to make cutters for each specific number of teeth, a series of “range cutters” was developed for use over a “range” of tooth numbers. Some profile accuracy was sacrificed but this was not a major problem at the loads and speeds being used at the time (pre-World War II) in most applications.

Form cutters are capable of producing internal and external spur gears, internal and external helical gears, and some types of straight bevel gears. The preferred production techniques for these gears have moved beyond form cutting, but the technique is still valid in some circumstances.

### Hobs

The term “hob” refers to a cutting tool with a spiral groove and radial slots or “gashes,” which serve as cutting surfaces (Fig. 2). Almost any uniformly spaced external tooth form can be hobbled, including non-symmetrical designs such as ratchets. Hob profiles do not “form” the tooth space but rather “generate” it as they pass through the blank. The design of a hob can be quite complex as the tool surface is the conjugate of the finished part.

“Standard” hobs are available in popular tooth forms and pressure angles for general purpose use. Higher production parts typically require custom hobs that may include modifications to provide profile modifications



Gear Cutting Tools, Fig. 1 Form Cutters

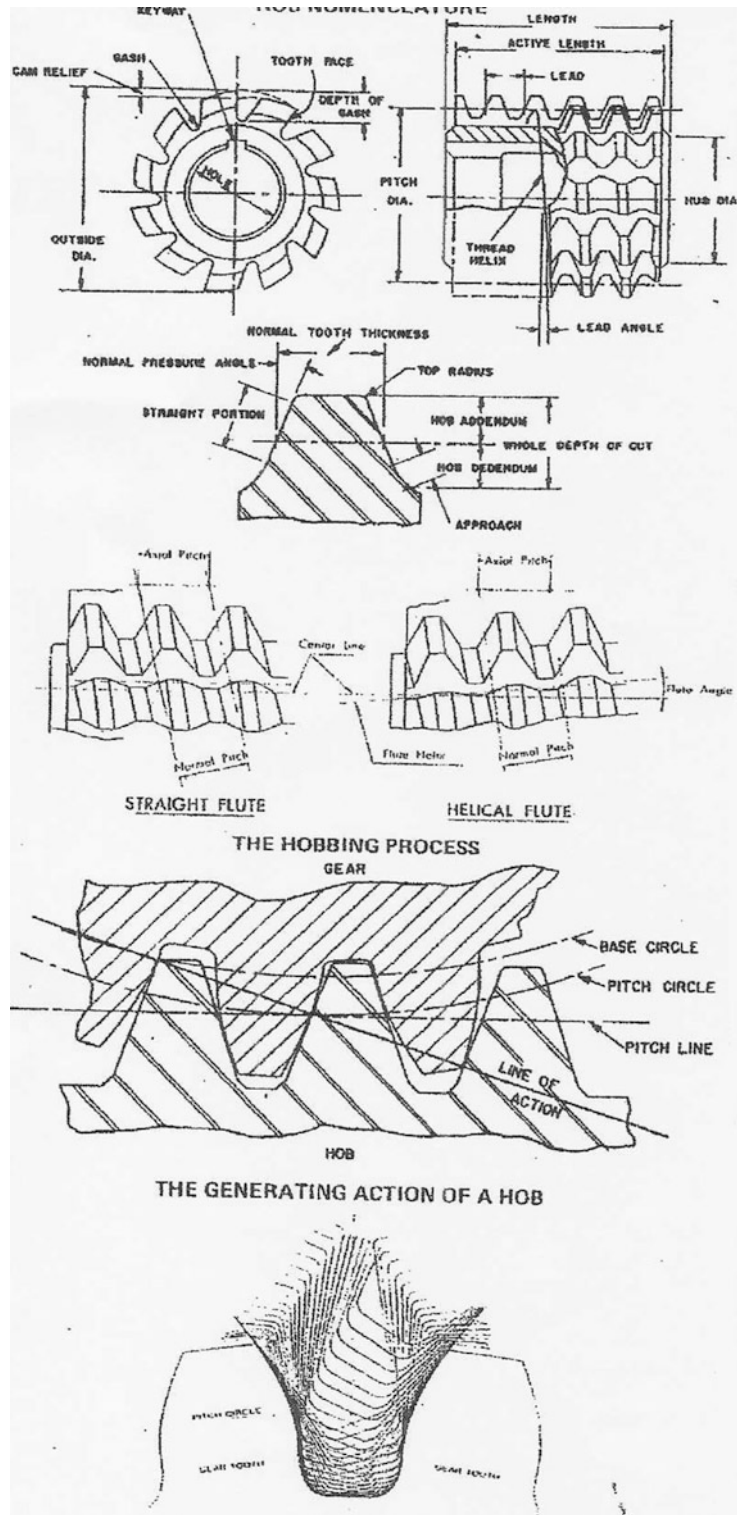
such as tip relief for quieter operation, protuberance to provide root clearance for subsequent finishing operations such as shaving, honing, or grinding, or changes to the tool pressure angle to alleviate the undercutting of the tooth.

High production hobs may have multiple treads or special gash geometry to permit very rapid cutting. If the teeth are to be finished after heat treat by gear grinding, multiple thread hobbing is especially attractive as the usual concerns over tooth form accuracy and flank surface finish are eliminated. Root fillet geometry and surface finish remain a concern for higher performance gears.

### Insert Cutters

Cutting tool steel has always been expensive, prompting tool designers to look for ways to use cheaper materials for parts of the tool that are not actually cutting the blank. A good example of this is found in bevel gear cutting, where the cutter may have an outside diameter of as much as 1 m (39.37”). The actual cutting surface on these tools is a high-speed steel or carbide “insert” that can be bolted into place in the cutter body. This allows the “wear surfaces” to be replaced when needed without throwing away the very large and expensive cutter body.

Insert technology has been applied to other types of gear cutting tools with varying degrees of success. It is extremely difficult, for example, to fasten inserts into

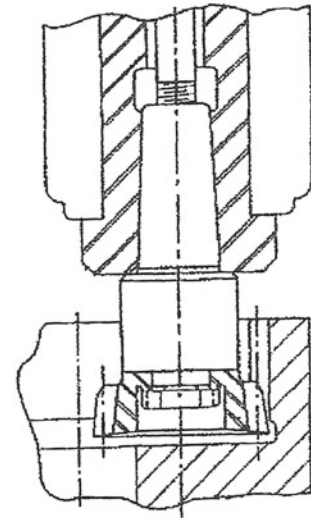
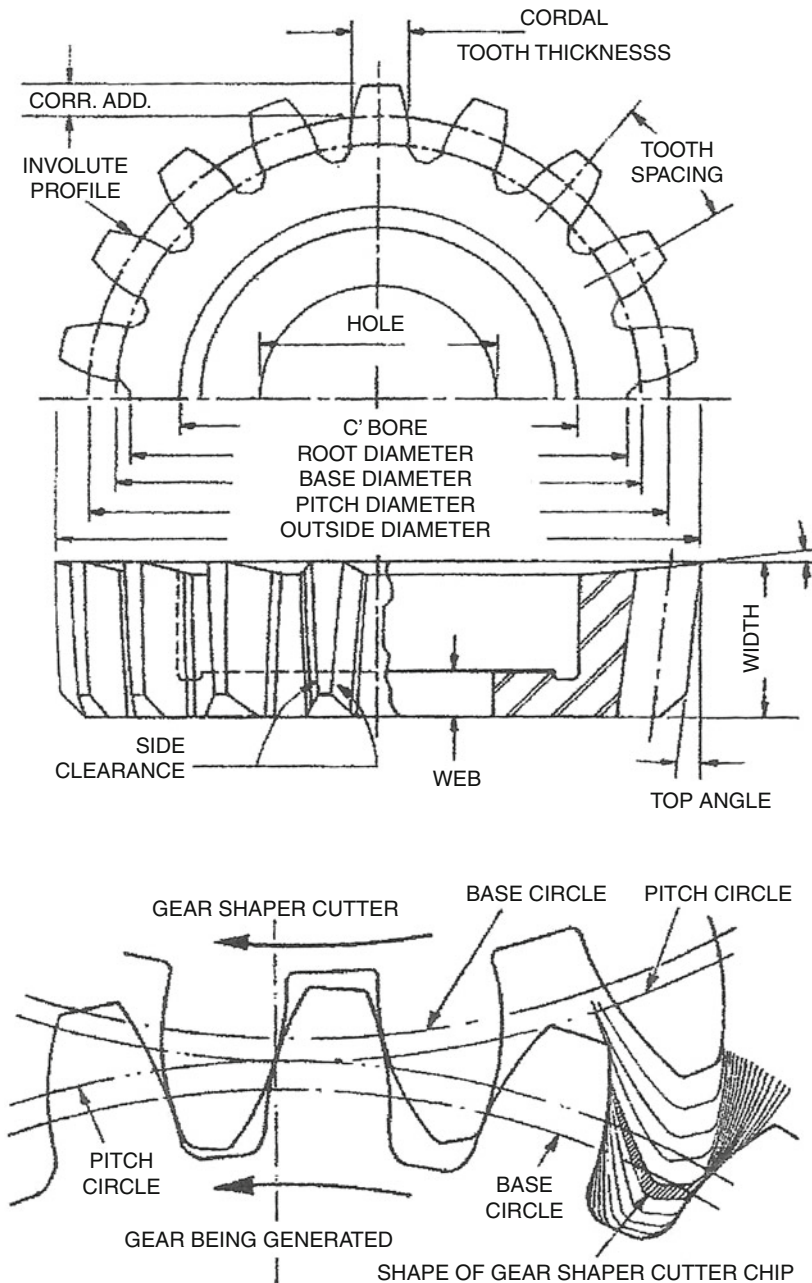


Gear Cutting Tools, Fig. 2 The generating action of a hob

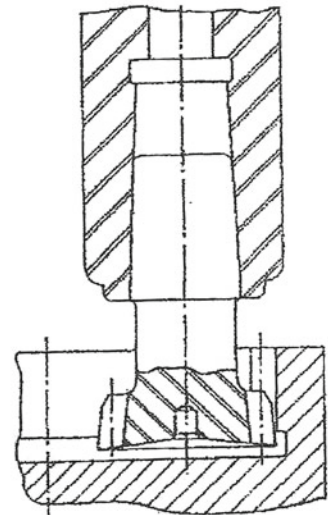


### Shaper Cutter Nomenclature

(Courtesy of Ash Gear and Supply Corporation)



Cutter & Holder



Shank Cutter

Gear Cutting Tools, Fig. 3 Shape of a gear shaper cutter chip

a small hob or milling cutter. Larger hobs have been made with multiple-toothed sections bolted in. Shaper cutters have been made with “throw away” wafers clamped in place to provide the cutting edges.

### Carbide Insert Cutters

Tungsten carbide offers the advantage of very high surface hardness (70 HRC or more) while maintaining useable toughness. The ability to use powder metal technology to



form carbide tool inserts made it practical to use carbide inserts rather than tool steel inserts in gear-cutting tools. The inserts may be custom shapes made to provide the desired tooth flank geometry or, in the case of gashing cutters, standard round, triangular, or diamond-shaped inserts arranged in the cutter body to provide a tooth flank very close to the desired pre-grind geometry. Gashing cutters may have 64 or more inserts installed in them.

In addition to high metal removal rates, carbide insert cutters do not require a special tool sharpening operation. If a cutting edge becomes dull, the insert can be indexed to a sharp edge or replaced while the cutter is still mounted in the machine. Standard triangular inserts, for example, have (6) different cutting edges that can be used; a diamond-shaped insert has (8). A full round insert, as might be used to generate a root fillet, can be rotated to provide numerous sharp cutting edges before it must be discarded.

### Shaper Cutters

Shaping is a generating process, like hobbing, but the cutting surfaces are curved rather than straight (Fig. 3). The shaper cutter is, in effect, a pinion, in mesh with the blank. As it advances along the face of the blank, it rotates in a timed relationship. At the end of each stroke, the cutter retracts slightly and reciprocates back to its starting position to begin another cutting stroke. Shaping can be used to produce internal and external spur teeth, internal and external helical gear teeth, and racks. While the reciprocating nature of the action means that as much as one half of the cutting cycle is not making chips, it also permits cutting teeth in parts a rotary motion cutter, such as a hob or mill, cannot. A good example of this is the “herringbone” tooth form where little or no gap is needed between the opposite helix hands.

### Rack Shaper Cutters

Rack shaper cutters can be thought of as a single row of cutting surfaces from a hob being used as a shaper cutter. The rack is, like a shaper cutter, in mesh with the blank but only the blank rotates. Rack shapers use a reciprocating motion and enjoy many of the advantages of a shaper while having a more easily fabricated and sharpened tool.

### Broaches

Broaches are used to make internal gears and splines. The tool consists of a series of cutting edges, each more closely resembling the tooth spaces of the part being made. Some parts may require multiple broaches to achieve the required configuration. The broach, or broaches, are typically pulled through the blank while it

is held in a specially designed fixture. Smaller broaches may be pushed through the blank.

### Coated Cutters

The introduction of titanium nitride and similar wear coatings in the early 1980s revolutionized the gear manufacturing business. Cutting tool lives were greatly increased overnight and the demand for new tools was reduced almost overnight. Tools have their greatest capability when freshly coated but retain some advantage even if sharpened between re-coatings.

### Cross-References

- [Gear Manufacturing](#)
- [Gear Manufacturing Machines](#)
- [Gear Surface Treatment](#)

### References

- D.W. Dudley, *The Evolution of the Gear Art*. American Gear Manufacturers Association Library of Congress Catalog Card Number: 72-78509
- D.W. Dudley, *Practical Gear Design*. McGraw-Hill Library of Congress Catalog Card Number: 53-11476
- D.W. Dudley, *Gear Handbook*. McGraw-Hill Library of Congress Catalog Card Number: 61-7304
- C.D. Schultz, *An Introduction to Gear Design*. Privately published 1987-1997. Available for free download at [www.beytagear.com](http://www.beytagear.com)

---

## Gear Dynamics

ROBERT ERRICHELLO

GEARTECH, Townsend, MT, USA

### Synonyms

[Dynamic factor](#); [Gear noise identification](#); [Gear vibration](#)

### Definition

Gear dynamics is the field of study concerned with gear vibration, noise, and gear tooth dynamic forces. It includes the application of geometry, metrology, kinematics, kinetics, solid mechanics, and acoustics to gear tooth excitation and dynamic response of a gear system.

### Scientific Fundamentals

#### System Dynamics

The dynamic response of a geared system depends on the distribution of mass, stiffness, and damping of the prime

mover, gearbox, and driven equipment, including shafts, couplings, and foundation. It is the system designer's responsibility to ensure that no natural frequencies exist near operating speeds. Natural frequencies, mode shapes, and dynamic response of the system to excitation forces must be calculated to ensure that the system is free of resonance.

### Gear Pair Resonance

It is possible for non-conjugate meshing action of gear teeth to excite a vibration between the pinion and gear if the frequency of the excitation matches the natural frequency of the gear pair. The gear pair spring/mass system consists of the pinion and gear masses connected by a spring representing the gear mesh stiffness. It is important to avoid operation near the resonant speed to avoid large gear tooth forces.

### Gear Tooth Excitation

In 1953, Strauch (1953) suggested that the meshing action of gear teeth induces a periodic excitation that creates continuous vibration that may build up to cause large dynamic forces on the gear teeth. Hence, for the first time, it was recognized that the excitation and dynamic response of a gear system is a steady-state vibration rather than a transient impact as previously thought. Zeman (1957) analyzed the effects of idealized periodic gear tooth profile deviations as a form of excitation. Other investigators realized periodic effects were important, for example, Utagawa (1958), Johnson (1958), and Kohler (1959) all contributed to the understanding of gear dynamics. However, it was most elegantly shown by Harris (1958) that transmission error (TE) is the best description of the excitation of a gear system. Harris defined TE as the departure from a constant velocity ratio between the input and output rotation of a pair of gears. He explained that TE is caused by displacements of gear teeth from their theoretically correct positions due to design, elastic deflections, stiffness variations, and manufacturing deviations.

Harris showed both analytically and experimentally that gear teeth could be designed for only one load because departures from the design load change the deflection and increase TE.

Figure 1 shows the principle of a TE tester. As the driving gear rotates the driven gear, the optical encoders sense the angular positions of each gear, and the signals are combined and analyzed to calculate the angular position of the driven gear relative to the driving gear. The TE is expressed as an angular displacement in (1) or a linear

displacement along the line of action in (2). Figure 2 shows an example TE graph. See AGMA 915 (2002) for definitions of the symbols used in Fig. 2.

$$TE = \theta_2 - \frac{z_1}{z_2} \theta_1 \quad (1)$$

$$TE = R_{b2} \left( \theta_2 - \frac{z_1}{z_2} \theta_1 \right) \quad (2)$$

Where:

TE	Transmission error from (1), rad
TE	Transmission error from (2), mm
$\theta_1$	Angular displacement of driving gear, rad
$\theta_2$	Angular displacement of driven gear, rad
$z_1$	Number of teeth in driving gear, dimensionless
$z_2$	Number of teeth in driven gear, dimensionless
$R_{b2}$	Base radius of driven gear, mm

### Manufactured Transmission Error (MTE)

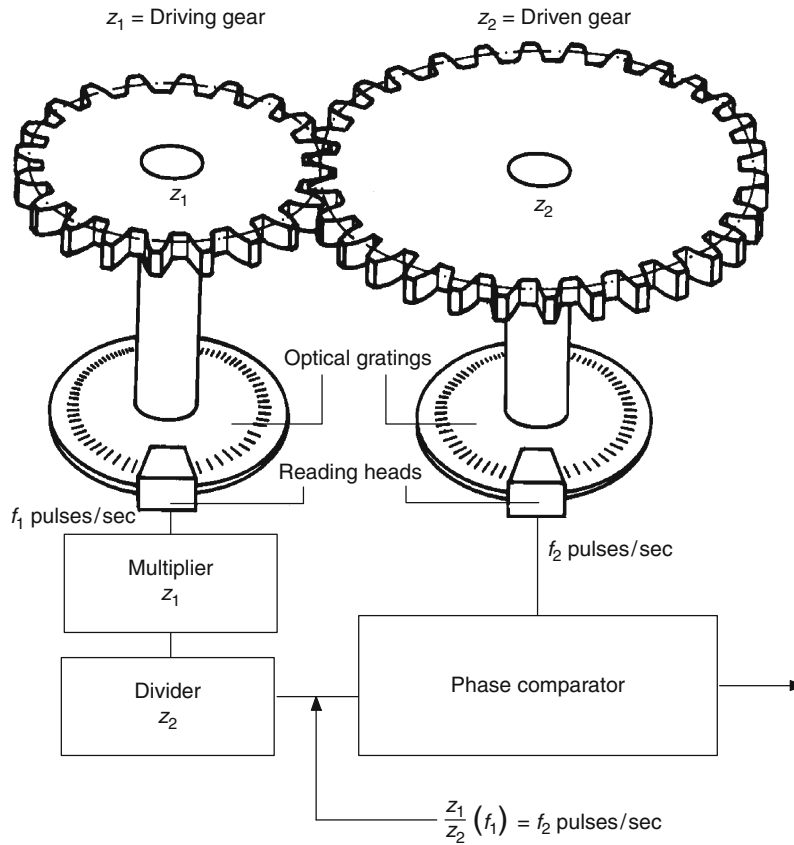
MTE is measured at low angular velocity with gear tooth contact maintained under light load. The MTE graph reflects TE as influenced by gear tooth manufacturing deviations including the combined elemental deviations of profile, helix, and pitch of both gears. MTE does not include the influence of load-induced gear tooth deflections and mesh stiffness variations. MTE is used for:

- Quality control of manufacturing
- Precise control of indexing accuracy
- Matching of gear pairs to minimize MTE

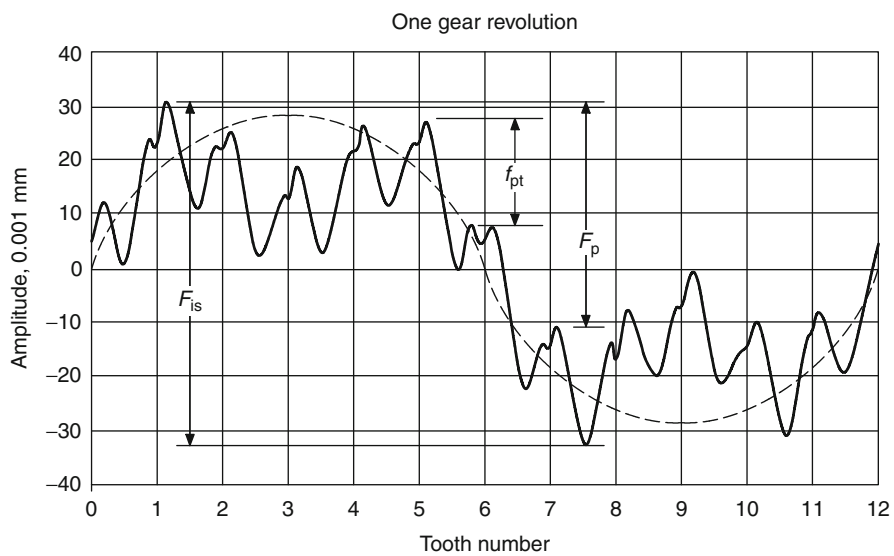
### Loaded Transmission Error (LTE)

LTE is measured under the design load and at low angular velocity. The LTE graph reflects the MTE (gear tooth deviations including the combined elemental deviations of profile, helix, and pitch of both gears) and includes the influence of load-induced gear tooth deflections and mesh stiffness variations. LTE is used for:

- Quality control of manufacturing
- Precise control of indexing accuracy
- Matching of gear pairs to minimize LTE
- Determining LTE excitation for use in analyses of gear tooth dynamic load and gear noise generation



Gear Dynamics, Fig. 1 Principle of a transmission error tester



Gear Dynamics, Fig. 2 Example transmission error graph

## Gear Noise

TE is the single most important factor in the generation of gear noise (Houser 1992). Several methods may be used to reduce gear noise:

- Reduce TE excitation
- Modify dynamic force paths and reduce vibrations between the gear mesh and the gear housing
- Reduce the gear housing's acoustic radiation efficiency
- Modify the environment surrounding the gearbox

## Dynamic Factor for Gear Rating

Industrial standards (ANSI/AGMA 2101-D04 2004; ISO 6336-1:1999 1999) use a dynamic factor,  $K_v$ , to account for internally generated gear tooth loads that are generated by TE. The dynamic factor relates the total dynamic tooth load to the static transmitted load with the following equation:

$$K_v = \frac{F_d + F_t}{F_t} \quad (3)$$

Where:

$K_v$	Dynamic factor, dimensionless
$F_d$	Incremental dynamic load not including the static transmitted load, N
$F_t$	Static transmitted load, N

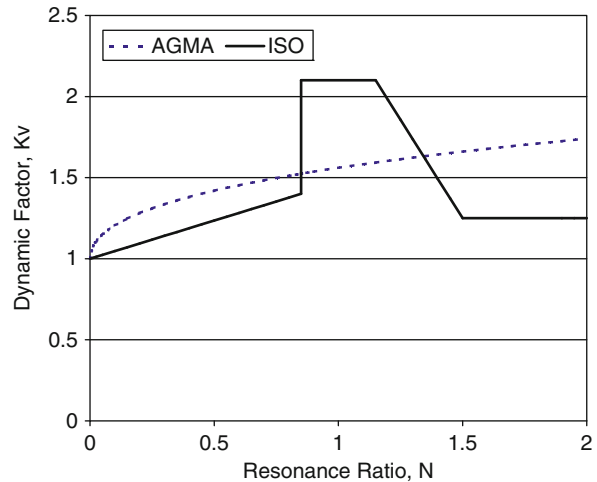
AGMA 2101 (2004) uses empirical equations for calculating  $K_v$ , and it explicitly excludes gear pair resonance. ISO 6336 (1999) considers gear pair resonance by modeling the gearset with a single-degree-of-freedom (SDOF) model consisting of the pinion and gear masses connected by a spring representing the gear mesh stiffness.

Figure 3 compares the AGMA and ISO dynamic factors for a particular gearset. The dynamic factor,  $K_v$ , is plotted against the resonance ratio,  $N$ , where  $N$  is the ratio of operating speed to the gear pair resonant speed. The ISO curve is divided into four sectors:

Sector	$N$
Subcritical	$0 \leq 0.85$
Main resonance	$>0.85 \leq 1.15$
Intermediate	$>1.15 \leq 1.5$
Supercritical	$>1.5$

## Overload Factor for Gear Rating

Industrial standards (ANSI/AGMA 2101-D04 2004; ISO 6336-1:1999 1999) use an overload factor,  $K_o$ , to



**Gear Dynamics, Fig. 3** AGMA and ISO factors for a particular gearset

account for vibratory loads that are superimposed on the mean operating load due to system dynamics including transient loads associated with startup, speed changes, overspeeds, shutdown, changes in process loads, variations in system operation, and system vibrations.

## Key Applications

Industrial standards such as (ANSI/AGMA 2101-D04 2004; ISO 6336-1:1999 1999) provide fundamental rating formulas to evaluate gear load capacity. These standards prescribe methods to determine the dynamic factor,  $K_v$ , and the overload factor,  $K_o$ .

## Acknowledgment

Figures 1 and 2 were extracted from AGMA 915-1-A02, *Inspection Practices- Part 1: Cylindrical Gears- Tangential Measurements*, with the permission of the publisher, the American Gear Manufacturers Association, 1500 King Street, Suite 201, Alexandria, Virginia 22314.

## Cross-References

- [Gear Tooth Modifications](#)
- [Gear Noise](#)
- [Involute Gear Profiles](#)

## References

- AGMA 915-1-A02, *Inspection Practices: Part 1. Cylindrical Gears- Tangential Measurements* (AGMA, Alexandria, VA, 2002)
- ANSI/AGMA 2101-D04, *AGMA Standard – Fundamental Rating Factors and Calculation Methods for Involute Spur and Helical Gear Teeth* (AGMA, Alexandria, 2004)

- S.L. Harris, Dynamic loads on the teeth of spur gears. *Proc. Inst. Mech. Eng.* **172**, 87–112 (1958)
- D. Houser, Gear noise, in *Dudley's Gear Handbook*, ed. by D.P. Townsend, 2nd edn. (McGraw-Hill, New York, NY, 1992). Chap. 14
- ISO 6336-1:1999, *ISO International Standard – Calculation of Load Capacity of Spur and Helical Gears: Part 1. Basic Principles, Introduction and General Influence Factors* (ISO, Geneva, Switzerland, 1999)
- D.C. Johnson, Excitation of resonant vibrations by gear tooth meshing effects, in *Proceedings of the International Conference on Gearing, Institute of Mechanical Engineers, London*, 1958, pp. 18–23
- H.K. Kohler, The mechanism and measurement of dynamic loading in spur gears, PhD thesis, University of Sheffield, England, 1959
- H. Strauch, “Zahnradsschwingungen” (gear vibrations). *Z. Ver. Dtsch. Ing.* **95**(6), 159–163 (1953)
- M. Utagawa, Dynamic load on spur gear teeth. *Bull. JSME* **1**(4), 397–403 (1958)
- J. Zeman, Dynamische Zusatzkräfte in Zahnradtrieben. *Z. Ver. Dtsch. Ing.* **99**(6), 244–254 (1957)

## Gear Dynamics and Noise

### ► Gear Noise

## Gear Efficiency

DENNIS A. LAUER

Kluber Lubrication North America L.P., The Villages,  
FL, USA

### Synonyms

Gearing energy losses; Gearing friction; PAG gear oils and efficiency; Polyglycol gear oils and efficiency; Synthetic gear oils and efficiency; Worm gear lubrication and efficiency

### Definition

Gear efficiency varies with the type of gear and the type of lubricant. In general, the advantage of synthetic gear oil over mineral gear oil varies with the type of gear system; [Table 1](#) shows typical results. Since the most dramatic differences are seen in worm drives, this discussion presents data obtained with a worm gearbox test rig.

Worm gears are often used as a simple and effective solution for power transmission applications. The latest calculation methods, new materials, modern manufacturing procedures, and synthetic lubricants

**Gear Efficiency, Table 1** Potential reduction of gearing losses and improvement of efficiency if synthetic gear oil is used instead of a mineral oil

Types of gear effect	Worm gears and hypoid gears	Spur gears and bevel gears with axis not offset
Reduction of total losses	30% and more	20% and more
Improved efficiency	15% and more	Up to 1%
Reduction of operating (steady-state) temperature	20°C and more	up to 12°C

have helped increase the load capacity of worm drive gearboxes over the last few years.

Typically, the development of a new gear oil starts with basic investigations modeling the tribological contact. Test rigs such as the four-ball tester and the oscillation-friction and wear tester are helpful, but the most effective method of testing gear oil is in a real gearbox, especially in terms of contact pressure and temperature. This is the only way to verify that the gear oil provides:

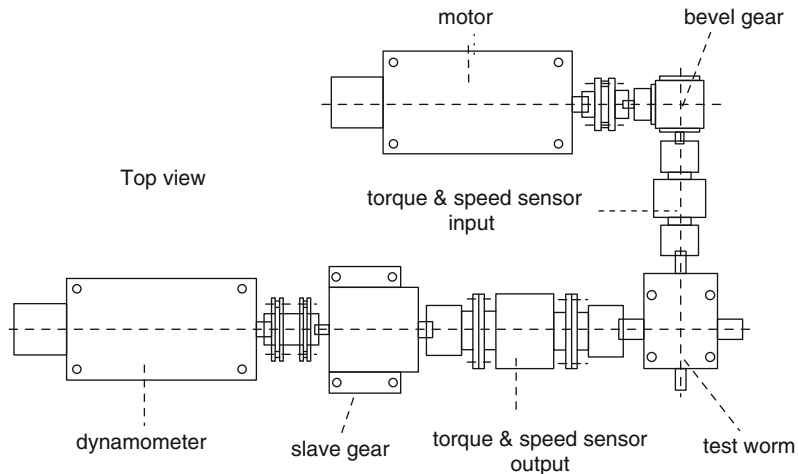
- Low friction and high efficiency
- Low operating temperatures to improve oil service life
- Low wear rate and increased fatigue resistance of materials

## Scientific Fundamentals

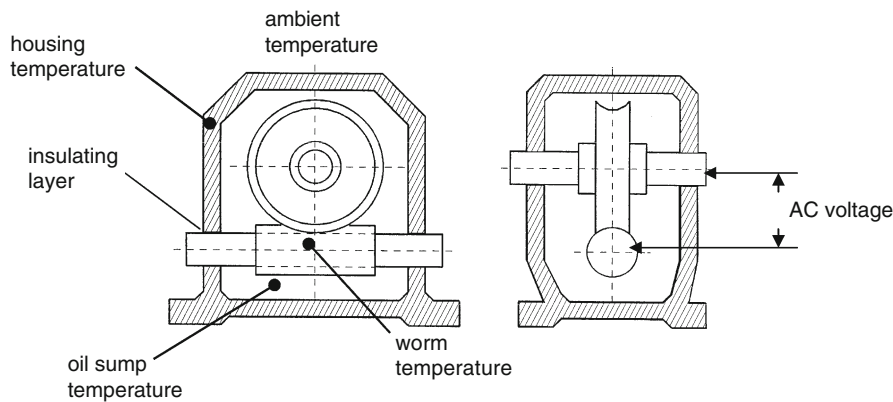
### Worm Gear Test Rig

The setup of the test rig ([Fig. 1](#)) makes it possible to measure speed and torque of the worm input and the output torque of the wheel ([Mann 1999](#)). The overall efficiency of the gearbox is calculated from measured values of input and output torque.

The test rig provides the opportunity to determine the lubricating regime in the mesh ([Fig. 2](#)) through a continuous operating method that determines changes in lubrication regime during the test run. With the worm shaft insulated from the gear box housing and an electrical potential established between the wheel and the worm, continuity can be measured when there is metal-to-metal contact. Full continuity implies boundary lubrication regime and as the continuity strength degrades to zero, the contact is passing through the mixed friction regime to full film lubrication.



**Gear Efficiency, Fig. 1** Test rig (Courtesy of Klüber Lubrication München KG)



**Gear Efficiency, Fig. 2** Determination of temperature and lubrication regime (Courtesy of Klüber Lubrication München KG)

The wear of the wheel is determined using the following two methods:

- Weight loss of the wheel after completion of the test run.
- Abrasion of the wheel flank while running. The abrasion amount is based on a continuous measurement of the wheel tooth thickness, which can be assessed whether it is a result of running in or of the actual operating conditions. The position of the worm shaft is established relative to the wheel shaft. Angular change from the starting position is monitored with electronic sensors and is used to calculate the reduction in tooth thickness, which is wear.

Different temperature measurements are part of the standard test.

- Bulk temperature of worm shaft
- Oil sump temperature
- Housing temperature
- Ambient temperature

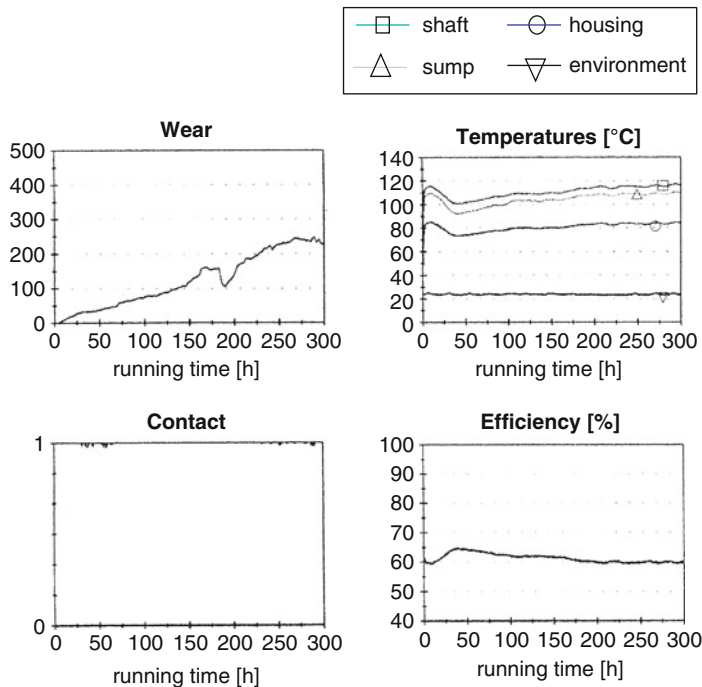
### Test Oils

**Table 2** illustrates data of the tested lubricants. A variety of different synthetic oils were chosen and compared with a mineral oil, which is often used for lubricating all types of gearboxes including worm gear drives.



Gear Efficiency, Table 2 Test oils

Oil	Base oil	Viscosity index	Notes (ISO grade and FZG load stage)
M460	Paraffinic mineral	95	CLP (A/8.3/90) > 12
PG460-1	Polyglycol	>220	CLP (A/16.6/140) > 12
PG460-2	Polyglycol	>220	CLP, "food grade lubricant" (A/16.6/90) > 12
PAO460-1	Polyalphaolefin	>150	CLP A/16.6/140) > 12
PAO460-2	Polyalphaolefin	>160	CLP, "food grade lubricant (A/8.3/90) > 12
E460	Ester	>150	CLP, "rapidly biodegradable" (A/8.3/90) > 12

**Test gear:**

Flender CUW 63  
 $i = 1 : 39$ ;  $a = 63$  mm  
 worm shaft 16MnCrS5 no. 11  
 worm gear GC-CuSn12Ni no. 11a

**Test conditions:**Running-in:

input speed: 350 1/min  
 output torque: 24 hrs 100 Nm  
 24 hrs 200 Nm  
 2 hrs 300 Nm

Permanent running:

input speed: 350 1/min  
 output torque: 300 Nm

Worm gear wear:

weight loss: 5.54 g  
 wear rate: 0.91  $\mu\text{m/h}$

Gear Efficiency, Fig. 3 Test results for mineral oil M460 (Courtesy of Klüber Lubrication München KG)

The viscosity of each oil is 460  $\text{mm}^2/\text{s}$  at 40°C (ISO VG 460). All tested oils are fully formulated (R&O, AW, and EP packages) and meet or exceed the requirements for CLP oils. Each oil used in the test passes the FZG scuffing test with load stage greater than 12.

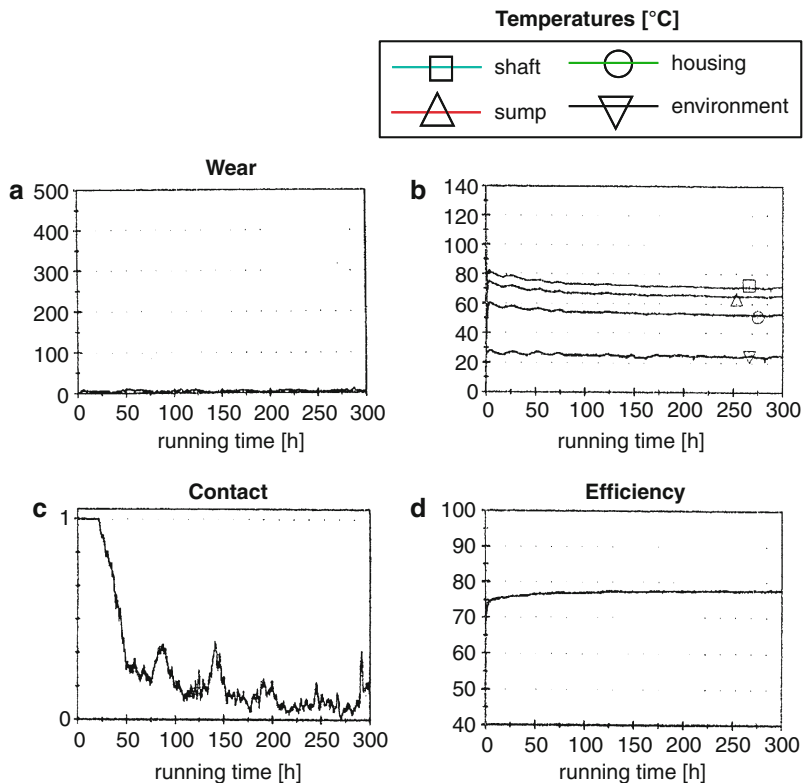
The M460 oil is based on a paraffinic mineral oil and contains a sulfur/phosphorous additive package. Because it is widely used in common gearbox applications, M460 oil was selected as the reference oil.

The polyglycol PG460-1 was selected as a typical modern high-performance gear oil. The ratio of ethylene oxide

to propylene oxide is 1:1. Because these types of base oils offer low friction coefficients, a low oil temperature and high efficiency can be expected, especially for gearboxes with a high sliding velocity in the mesh.

PG460-2, a polyglycol, was designed as a "food grade lubricant." The "guidelines of Sec. 21 CFR 178.3570 of the FDA" are met by the formulation and the base oil properties are comparable to PG460-1. Only raw materials meeting the food grade specifications are used.

PAO460-1 is a polyalphaolefin oil that contains a sulfur/phosphorous additive package and was originally

**Test gear:**

Flender CUW 63

 $i = 1 : 39$ ;  $a = 63$  mm

worm shaft 16MnCrS5 no. 17

worm gear GC - CuSn12Ni no. 17

**Test conditions:**Running-in:

input speed: 350 1/min

output torque: 24 hrs 100 Nm

24 hrs 200 Nm

2 hrs 300 Nm

**Permanent running:**

input speed: 350 1/min

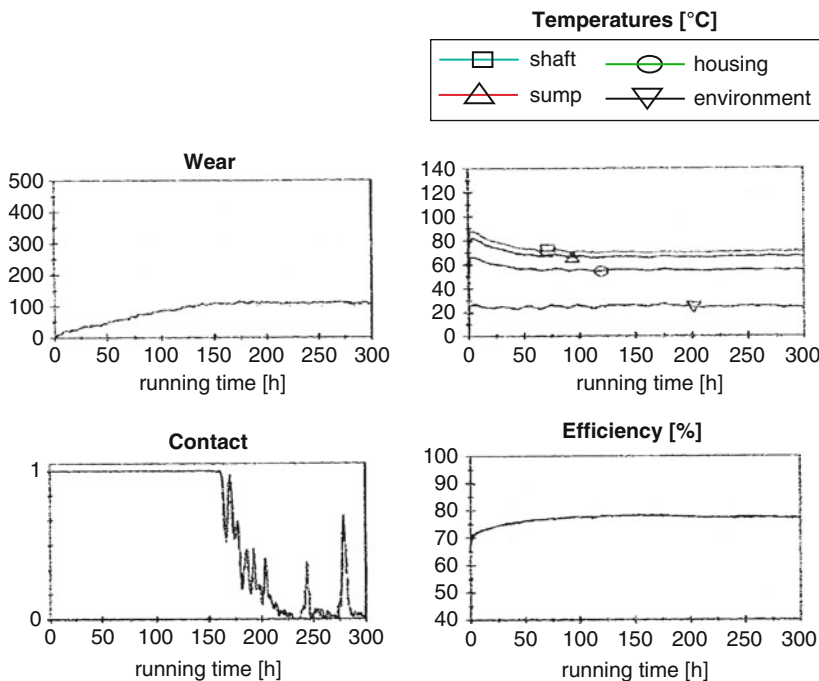
output torque: 300 Nm

**Worm gear wear:**

weight loss: 0.22 g

wear rate: ---  $\mu\text{m/h}$ 

Gear Efficiency, Fig. 4 Test results for polyglycol PG460-1 (Courtesy of Klüber Lubrication München KG)

**Test gear:**

Flender CUW 63

 $i = 1 : 39$ ;  $a = 63$  mm

worm shaft 16MnCrS5 no. 37

worm gear GC - CuSn12Ni no. 37

**Test conditions:**Running-in:

input speed: 350 1/min

output torque: 24 hrs 100 Nm

24 hrs 200 Nm

2 hrs 300 Nm

**Permanent running:**

input speed: 350 1/min

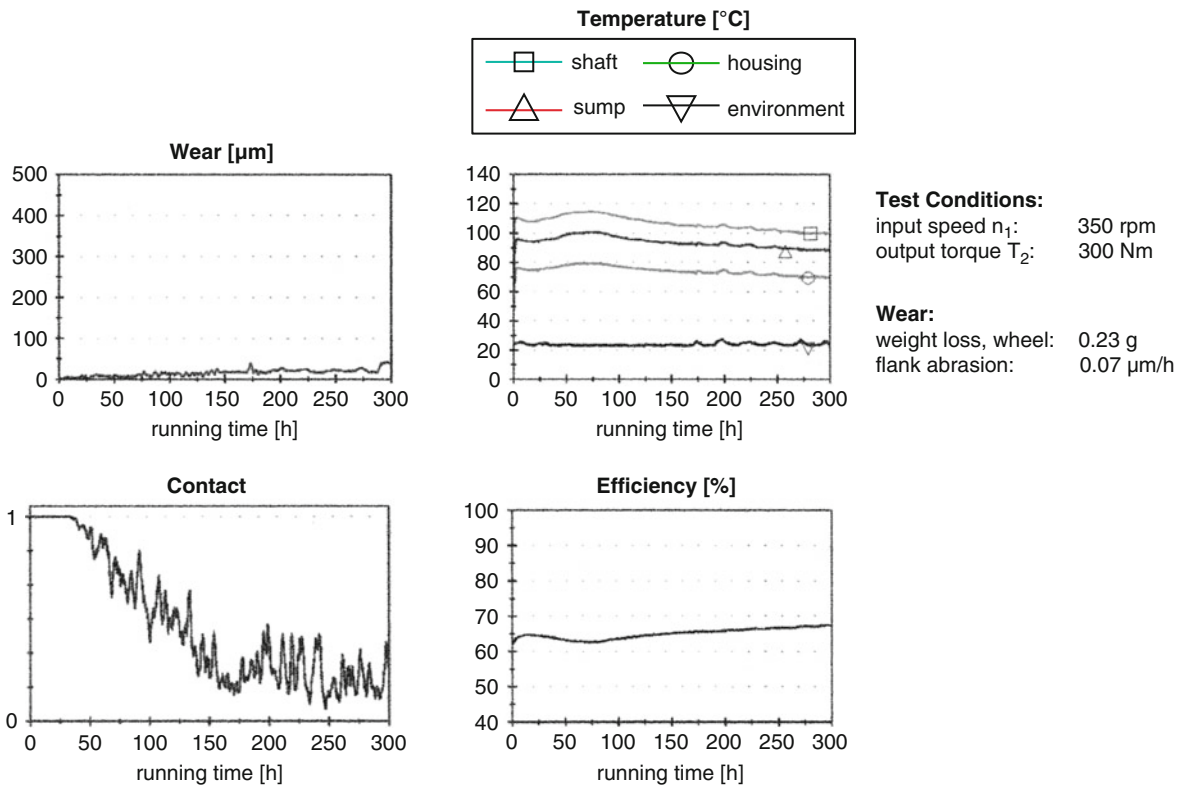
output torque: 300 Nm

**Worm gear wear:**

weight loss: 3.22 g

wear rate: 0.117  $\mu\text{m/h}$ 

Gear Efficiency, Fig. 5 Test results for polyglycol PG460-2 (Courtesy of Klüber Lubrication München KG)



Gear Efficiency, Fig. 6 Test results for PAO460-1 (Courtesy of Klüber Lubrication München KG)

developed for application in spur, helical and bevel gears. Polyalphaolefin oils also have lower friction coefficients than mineral oils.

The oil PAO460-2 is a polyalphaolefin and contains base oils and additives listed in the aforementioned FDA list. Acceptable for use in the food and pharmaceutical industry, it is classified as a “food grade lubricant.”

Intended for use wherever biodegradability is required, E460 oil is rapidly biodegradable and based on a synthetic ester. The CEC-L-33-A-94 standard states that oil is rapidly biodegradable when more than 70% of the lubricant is biodegradable within 21 days.

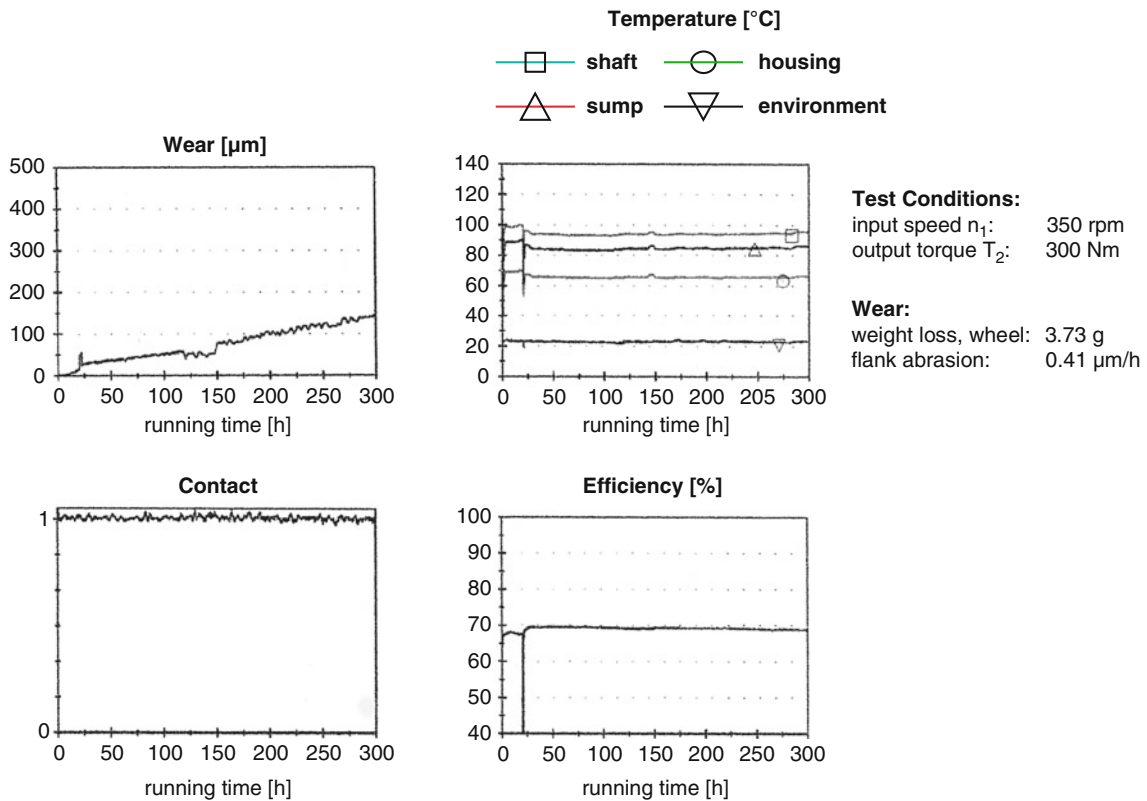
### Worm Gear Test Results

Figure 3 demonstrates the results of a test run with M460. Most measurements were evaluated after 250 h of testing. Test results show a shaft temperature of approximately 115°C (239°F) and the total efficiency of the gearbox is nearly 60%. Figure 3c identifies the lubrication regime. When the curve is on top of the diagram, mainly mixed friction appears in the contact zone while a curve at the bottom means better lubricating conditions and

a separating oil film. Although this method does not measure the exact film thickness, in the case of oil M460, the test shows mixed friction which leads to continuous wear of the wheel. The apparent decrease of wear at 200 h is actually due to a measuring error, not a weight gain. During running in and the test run, weight loss of the wheel is 5.5 g.

The PG460-1 (Fig. 4) test results show lower temperatures; after 150 h, the shaft temperature of the worm remains constant at 70°C (158°F) and the efficiency reaches 78%. Figure 4c gives an explanation of this high efficiency. After a period of approximately 50 h, with the gear running at full load, the film thickness increases and the lubrication conditions become better. High efficiency is achieved with a good lubricating regime and low frictional losses. The measured weight loss of the wheel is very low and appears to be a result of running in wear. Figure 4a shows no reduction of teeth thickness during the entire test run.

The measurements for PG460-2 (Fig. 5) show interesting results. While measured temperatures and efficiency are comparable to the results for PG460-1,



**Gear Efficiency, Fig. 7** Test results for PAO460-2 (Courtesy of Klüber Lubrication München KG)

a clear run in effect is recognizable. Also, after finishing the running in procedure, the wear of the wheel continues until 150 h when the wear of the wheel stops. The size and orientation of the contact pattern accounts for the change in wear rate. When the contact pattern reaches its maximum area, a separating oil film is formed and this leads to maximum efficiency and minimum temperatures.

PAO460-1 has a shaft temperature of approximately 100°C with the measured efficiency becoming stable at approximately 67% and the wear rate is relatively low, 0.07 μm/h (Fig. 6). After approximately 50 h, a separating oil film occurs.

Similar shaft temperatures and efficiencies as PAO460-1 are seen with PAO460-2 (Fig. 7). Due to the lack of separating oil film the wear rate is higher, 0.41 μm/h. Apparently, the additives in the PAO460-2 oil do not form a protective layer that is sufficient to prevent wear on the wheel flanks.

Efficiency for the E460 is nearly 70%, the shaft temperature is around 100°C, and the wear rate is

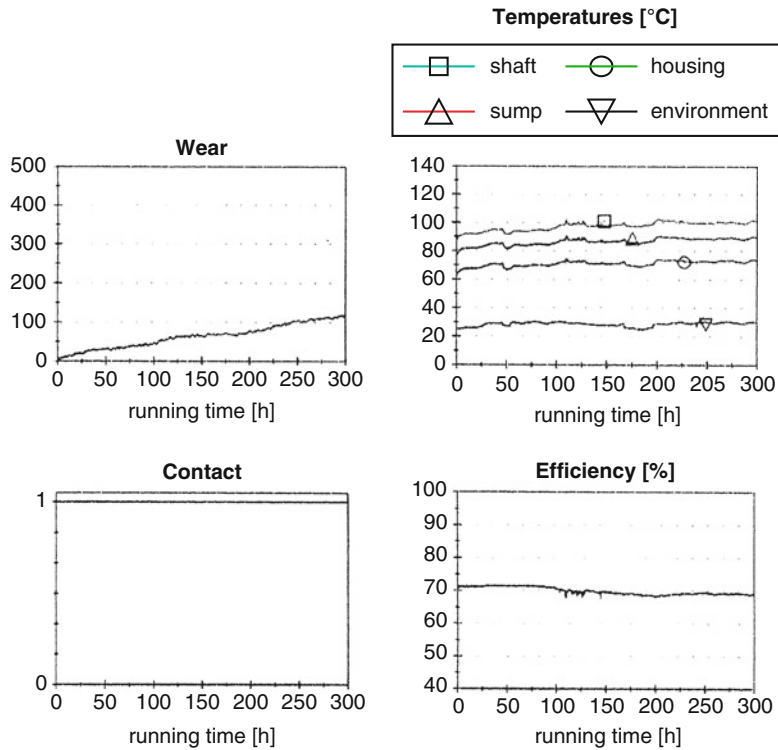
approximately 0.44 μm/h. As with PAO460-2, no separating lubricating oil film is formed under the selected test conditions (Fig. 8).

### Measured Wear

Figure 9 compares the measured wear rate for the three standard industrial oils tested. It is evident that the gears experienced the greatest wear rate with the mineral oil. As stated previously, the wear rate for the PAO and the ester oil were essentially the same, approximately half of that of mineral oil. The polyglycol shows the lowest wear rate at about 25% of the PAO wear rate.

### Measured Efficiency

A comparison of the efficiencies of these three test oils was also made (Fig. 10). It is intuitive that wear and efficiency would have an inverse relationship. Therefore, the mineral oil has the lowest efficiency at 60%, with the PAO and ester efficiency being 8 percentage points higher at approximately 68%. At approximately 18 percentage points higher than mineral oil, polyglycol has a value of 78%.

**Test gear:**

Flender CUW 63

 $i = 1 : 39$  ;  $a = 63$  mm

worm shaft 16MnCrS5 no. 36

worm gear GC - CuSn12Ni no. 36

**Test conditions:**Running-in:

input speed: 350 1/min

output torque: 24 hrs 100 Nm

24 hrs 200 Nm

2 hrs 300 Nm

Permanent running:

input speed: 350 1/min

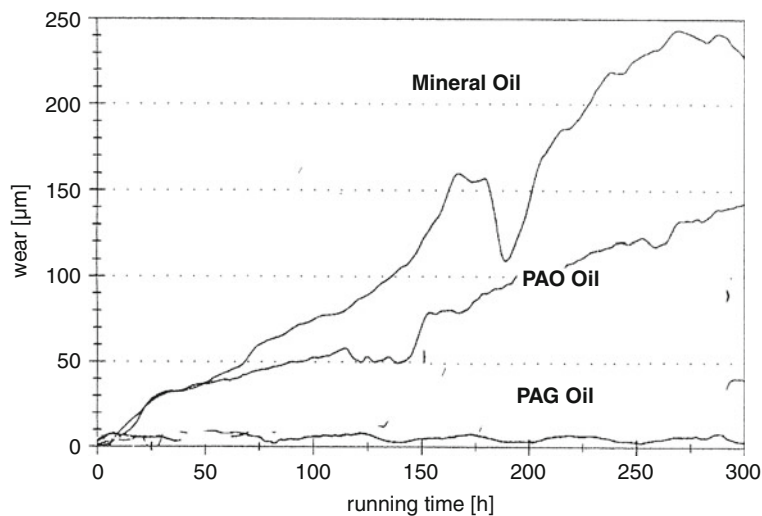
output torque: 300 Nm

**Worm gear wear:**

weight loss: 2.30 g

wear rate: 0.44  $\mu\text{m}/\text{h}$ 

Gear Efficiency, Fig. 8 Test results for E460 (Courtesy of Klüber Lubrication München KG)

**Test gear**

Flender CUW 63

 $i = 1 : 39$  ;  $a = 63$  mm

worm shaft 16 MnCrS5

worm gear GC - CuSn12Ni

**Test conditions**Running-in:

input speed: 350 1/min

output torque: 24 hrs 100 Nm

24 hrs 200 Nm

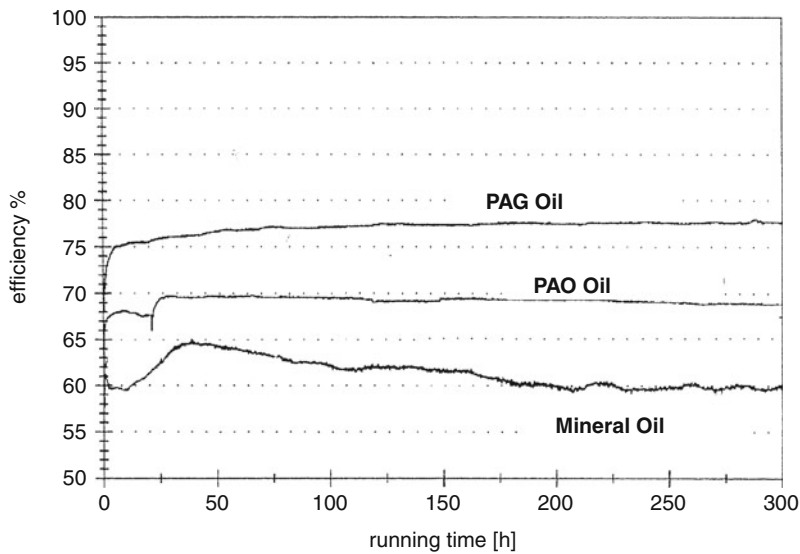
2 hrs 300 Nm

Permanent running

input speed: 350 1/min

output torque: 300 Nm

Gear Efficiency, Fig. 9 Wear profiles of three test oils (Courtesy of Klüber Lubrication München KG)

**Test gear:**

Flender CUW 63  
 $i = 1 : 39$  ;  $a = 63$  mm  
 worm shaft 16 MnCrS5  
 worm gear GC - CuSn12Ni

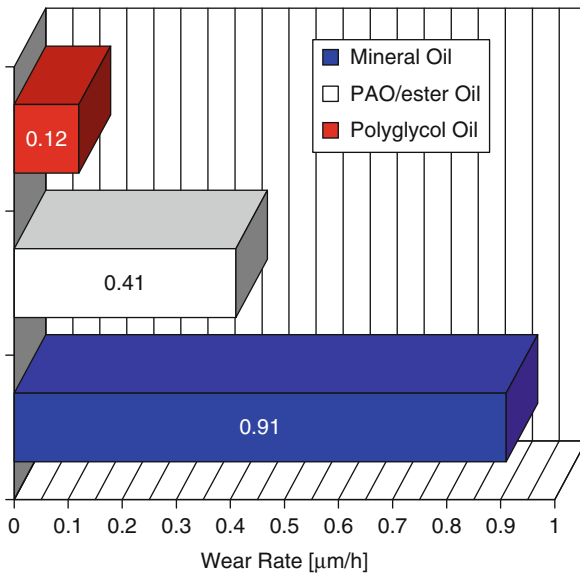
**Test conditions:**

Running-in:  
 input speed: 350 1/min  
 output torque: 24 hrs 100 Nm  
 24 hrs 200 Nm  
 2 hrs 300 Nm

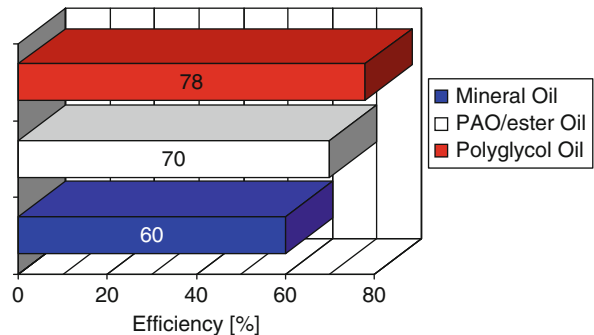
**Permanent running:**

input speed: 350 1/min  
 output torque: 300 Nm

**Gear Efficiency, Fig. 10** Efficiencies of three test oils (Courtesy of Klüber Lubrication München KG)



**Gear Efficiency, Fig. 11** Reduction of wear rate (Courtesy of Klüber Lubrication München KG)



**Gear Efficiency, Fig. 12** Increase in efficiency (Courtesy of Klüber Lubrication München KG)

influences the efficiency of a gear box. The influence of additives on wear is important for lubricating conditions with low film thickness, such as during running in and low speed operation.

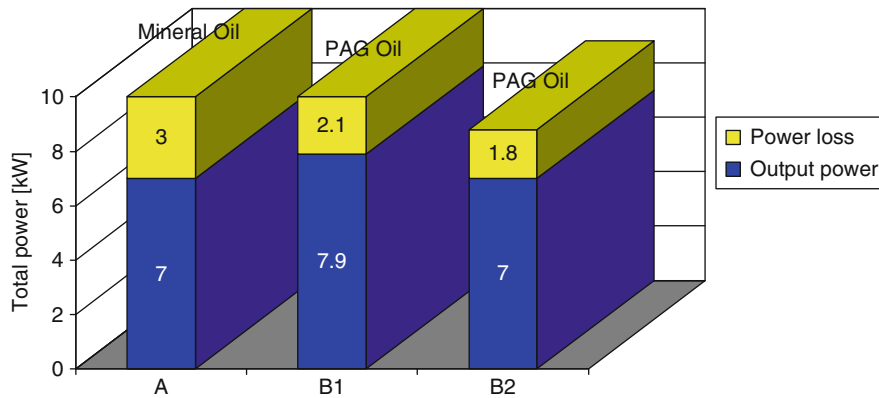
When examining the wear rate, the PAO and ester oil is about half of mineral oil and polyglycol is about one eighth the wear rate of mineral oil (Fig. 11).

Improved efficiency can be significant to the end user as well as the OEM (Fig. 12). End users will save energy that will reduce operational costs or operate at the same energy level with an increase in the output level and productivity, whereas for the OEM increased efficiency

## Key Applications

The results show that synthetic oils, especially polyglycol oils, with low friction coefficients improve efficiency and reduce temperatures. The base oil friction behavior





B1: 12.5 % increase of output power B2: 11.5 % reduction of energy costs

**Gear Efficiency, Fig. 13** Reduction in energy cost (Courtesy of Klüber Lubrication München KG)

can impact some design decisions. The gearbox output rating can be increased with the current design or the gearbox can be redesigned to be less costly and maintain the current output rating. A significant energy savings or power output advantage over mineral oil is provided by polyglycol gear oil (Fig. 13).

In order to increase fuel economy in automobiles, the use of low-viscosity lubricants is being emphasized. When conventional viscosity oils are replaced with lower viscosity oils, fuel mileage can increase as much as 8%. To protect the gears from wear due to the reduced  $\lambda$  ratio, the oils must contain more robust antiwear and antiscaff additives. In lieu of changing the additive package, the  $\lambda$  ratio can be increased by reducing the surface roughness. Superfinishing the surface using chemically accelerated vibratory finishing is a practical and proven approach for accomplishing this. Both laboratory and field testing have demonstrated that superfinished components exhibit lower friction, operating temperature, and wear, all of which translate into higher efficiency and lower fuel consumption (Winkelmann et al. 2009).

## Cross-References

- Friction in Gears
- Gear Lubricants
- Gear Lubrication
- Wear in Gears

## References

- U. Mann, Synthetic oils for worm gear lubrication. AGMA Technical Paper 99FTM17 (1999).
- L. Winkelmann, O. El Saeed, M. Bell, The capacity of superfinished vehicle components to increase fuel economy part 1. Gear Technol. 26(1), 50–60 (2009)

## Gear EHL Film Thickness and Wear Risk Analysis

ROBERT ERRICHELLO

GEARTECH, Townsend, MT, USA

## Synonyms

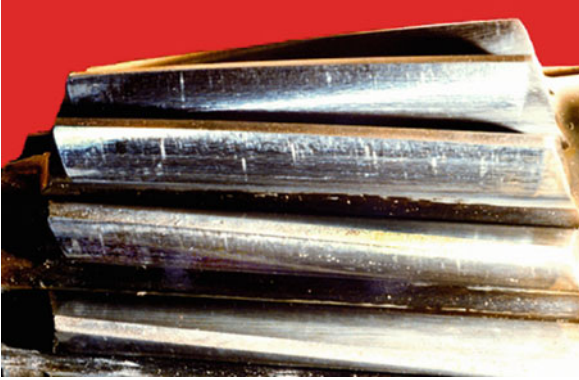
Break-in of gear surfaces; Mild adhesion; Moderate adhesion; Normal wear; Running-in of gears

## Definition

Adhesion is localized damage caused by solid-phase welding between surfaces in relative motion. It is accompanied by transfer of metal from one surface to another due to welding and subsequent tearing. ANSI/AGMA 1010-E95 (1995) describes the characteristics of the failure mode. AGMA 925-A03 (2003) gives information on effects of lubrication on gear surface distress. Figure 1 shows adhesive wear on a helical pinion.

## Scientific Fundamentals

Gear tooth load capacity based on the probability of wear is calculated according to industrial standards such as AGMA 925-A03 (2003), which calculates wear risk based on EHL specific film thickness using the Dowson and Toyoda equation (Dowson and Toyoda 1978). Although specific film thickness is a major factor governing adhesive wear, the properties of the lubricant and its additives exert a strong influence (Errichello 1992). Therefore, to help prevent adhesive wear, the physical and chemical properties of the lubricant should conform to the requirements



**Gear EHL Film Thickness and Wear Risk Analysis, Fig. 1**  
Adhesive wear on a helical pinion

of ANSI/AGMA 9005-D94 (1994). It is especially important that the lubricant have adequate antiwear additives to ensure adequate wear resistance.

### EHL Film Thickness Analysis

AGMA 925-A03 (2003) calculates EHL film thickness with equation (1). See reference (AGMA 925-A03 2003) for information on the variables used in (1) through (10).

$$h_c = H_c \rho_n \times 10^3 \quad (1)$$

$$H_c = 3.06 \frac{G^{0.56} U^{0.69}}{W^{0.10}} \quad (2)$$

$$G = \alpha E_r \quad (3)$$

$$U = \frac{\eta_M v_e}{2 E_r \rho_n} \times 10^{-6} \quad (4)$$

$$W = \frac{X_\Gamma w_n}{E_r \rho_n} \quad (5)$$

$$\lambda = \frac{h_c}{\sigma_x} \left( \frac{L_x}{2b_H} \right)^{1/2} \quad (6)$$

$$\sigma_x = (R_{a1}^2 + R_{a2}^2)^{1/2} \quad (7)$$

$$w_n = \frac{F_{wn} K_O K_H K_V}{L_{min}} \quad (8)$$

$$b_H = \left( \frac{8 X_\Gamma w_n \rho_n}{\pi E_r} \right)^{1/2} \quad (9)$$

$$E_r = 2 \left( \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \right)^{-1} \quad (10)$$

where

$h_c$	Central film thickness, $\mu\text{m}$
$H_c$	Central film thickness, dimensionless
$\rho_n$	Normal relative radius of curvature, mm
$G$	Materials parameter, dimensionless
$U$	Speed parameter, dimensionless
$W$	Load parameter, dimensionless
$\alpha$	Pressure-viscosity coefficient, $\text{mm}^2/\text{N}$
$E_r$	Reduced modulus of elasticity, $\text{N}/\text{mm}^2$
$\eta_M$	Dynamic viscosity at gear tooth temperature $\theta_M$ , $\text{mPa}\cdot\text{s}$
$v_e$	Entraining velocity, $\text{m}/\text{s}$
$X_\Gamma$	Load sharing factor, dimensionless
$w_n$	Normal unit load, $\text{N}/\text{mm}$
$\lambda$	Specific film thickness, dimensionless
$L_x$	Filter cutoff of wavelength $x$ , mm
$\sigma_x$	Composite surface roughness for filter cutoff wavelength, $L_x$ , $\mu\text{m}$
$b_H$	Semi-width of the Hertzian contact band, mm
$R_{a1}, R_{a2}$	Average surface roughness (pinion, gear) at $L_x$ , $\mu\text{m}$
$F_{wn}$	Total load normal to the gear tooth profiles at the operating pitch point, N
$K_O$	Overload factor accounts for externally applied loads induced by dynamics of the gear system, dimensionless
$K_H$	Load distribution factor accounts for non-uniform distribution of load along the lines of contact between mating gear teeth, dimensionless
$K_V$	Dynamic factor accounts for internally generated gear tooth loads induced by non-uniform meshing action of gear teeth, dimensionless
$L_{min}$	Minimum contact length, mm
$\nu_1, \nu_2$	Poisson's ratio (pinion, gear), dimensionless
$E_1, E_2$	Modulus of elasticity (pinion, gear), $\text{N}/\text{mm}^2$

### Wear Risk Analysis

AGMA 925-A03 (2003) uses empirical methods to relate calculated values of minimum specific film thickness,  $\lambda_{min}$ , to empirical values of mean specific film thickness,  $\mu_{\lambda,min}$ , with equation (11), which uses a Gaussian distribution to calculate the probability of wear. See reference (AGMA 925-A03 2003) for information on the variables used in (11).

$$x = \frac{\lambda_{min} - \mu_{\lambda,min}}{\sigma_{\lambda,min}} \quad (11)$$

where

$x$	Standard normal variable, dimensionless
$\lambda_{\min}$	Minimum specific film thickness, dimensionless
$\mu_{\lambda,\min}$	Mean minimum specific film thickness, dimensionless
$\sigma_{\lambda,\min}$	Standard deviation of mean minimum specific film thickness, dimensionless

## Key Applications

Industrial standards such as (AGMA 925-A03 2003) provide fundamental rating formulas to evaluate gear tooth load capacity. Application standards tailor the fundamental formulas to particular applications such as wind turbine gears (ANSI/AGMA/AWEA 6006-A03 2003) or marine gears (ANSI/AGMA 6032-A94 1994).

## Cross-References

- [EHL Film Thickness Behavior](#)
- [Gear Lubricants](#)
- [Gear Lubrication](#)

## References

- AGMA 925-A03, *AGMA Information Sheet – Effect of Lubrication on Gear Surface Distress* (AGMA, Alexandria, 2003)
- ANSI/AGMA 1010-E95, *AGMA Standard – Appearance of Gear Teeth – Terminology of Wear and Failure* (AGMA, Alexandria, 1995)
- ANSI/AGMA 6032-A94, *AGMA Standard – Standard for Marine Gear Units: Rating* (AGMA, Alexandria, 1994)
- ANSI/AGMA 9005-D94, *Industrial Gear Lubrication* (AGMA, Alexandria, 1994)
- ANSI/AGMA/AWEA 6006-A03, *AGMA Standard – Standard for Design and Specification of Gearboxes for Wind Turbines* (AGMA, Alexandria, 2003)
- D. Dowson, S. Toyoda, A central film thickness formula for elastohydrodynamic line contacts. In *5th Leeds-Lyon Symposium Proceedings*, Paper 11 (VII), 1978, pp. 60–65
- R. Errichello, Friction, Lubrication, and Wear of Gears, in *ASM Handbook*, vol. 18 (ASM International, Materials Park, 1992), pp. 535–545

## Gear Friction Reduction

- [Gear Lubrication](#)

## Gear Grinding

- [Gear Manufacturing](#)

## Gear Lubricants

BRIAN M. O'CONNOR

The Lubrizol Corporation, Wickliffe, OH, USA

## Synonyms

[EP oil or lubricant](#); [Gear oil](#)

## Definition

Gear lubricants are fluids formulated with specific chemical additives designed to protect surfaces that encounter high contact stresses and sliding speeds from catastrophic wear. These lubricants are generally used in a wide variety of applications, particularly when there is a high risk of adhesive wear (scuffing) or high loads are being transmitted by the gear system.

## Scientific Fundamentals

Gear lubrication covers a very wide range of applications, but it is generally divided into two main categories: automotive and industrial. Automotive applications would include drive axles and manual transmissions while industrial gear lubricants are usually referred to by the type of gearbox being used (i.e., enclosed or open). Regardless of the application, the gear lubricant is intended to protect the mating surfaces from excessive wear, minimize friction, and help to dissipate the heat generated in the contact. This is done through optimizing both the physical and chemical properties of the lubricant for the application. The physical properties are governed by the choice of base oil and thickener used in the formulation while the chemical properties come from a balanced selection of individual chemicals aimed at specific protection needs (i.e., wear, corrosion, oxidation, foaming, and so on). Both the automotive and industrial applications have similar general performance needs that might suggest a one-size-fits-all lubricant, but there are sufficient differences in the application characteristics that require a different approach for each.

The largest difference requiring more than one gear lubricant is the type of gearing to be protected. Final drive axles commonly use hypoid gearing. This type gearing provides quiet operation with a relatively high load capacity. Because of its design and kinematics the hypoid gear has a higher risk of scuffing (adhesive wear) than parallel axis or other simple bevel gearing and, therefore, must be formulated with higher levels of certain chemicals to minimize this risk. Industrial gearing, on the other hand, tends to use more parallel axis and spiral bevel gearing and the applications in many cases involve very large size gears operating at

Gear Lubricants, Table 1 General characteristics of base fluids

General class	Mineral	Synthetic			
Chemical type	Broad hydrocarbon distribution	Poly alpha olefin (PAO)	Ester	Poly alkylene glycol (PAG)	Phosphate esters
Viscosity – temperature relationship (typical viscosity Index (VI))	90–130	130–150	50–140	200–240	<100
Specific heat (relative)	1.0	1.3–1.5	1.1–1.3		1.0–1.2
Pressure-viscosity at 1 GPa (Relative)	1.0	0.8	0.5	~1.0	1.0–1.1
Compatibility with mineral fluids	Excellent	Good	Excellent	Poor	Good
Compatibility with PAO fluids	Good	Excellent	Excellent	Poor	Good
Additive compatibility	Good to excellent	Good	Good	Limited	Good

lower speeds. The combination of line contact (spur and helical gearing) and lower speeds pushes the emphasis on the physical properties of the lubricant, specifically toward higher viscosities. Smaller gearing such as those used in manual transmissions and many other industrial drive applications require less chemical additive due to the nature of the contact formed by parallel axis gears (line) versus the contact in a hypoid gear (elliptical).

In a broad sense, gear lubricants are composed of a minimum of two components, one for the physical aspects and the other for the chemical or performance aspects of the application. Additionally, there can be supplemental components used to further refine or enhance the basic properties. These supplemental components can include polymer-type thickeners, low-temperature flow control agents (pour point depressants), and friction-modifying chemical agents for special applications.

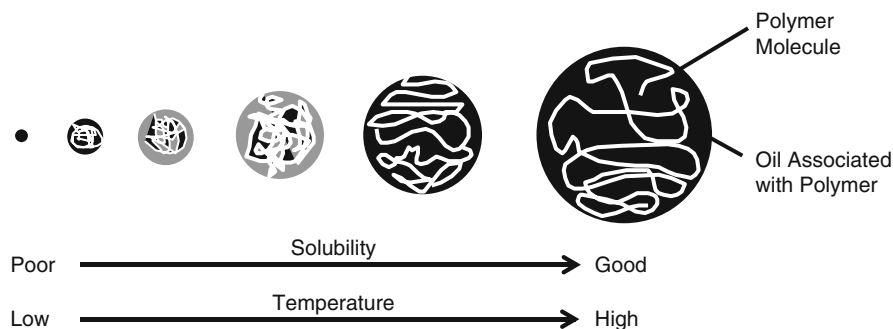
### Physical Properties

The physical properties of a lubricant, whether it is to be used for automotive or industrial application, are largely derived from the base stock(s) from which they are produced. For example, the crude source, the fraction or cut, and the amount of refining, such as dewaxing a given mineral oil, can significantly alter the way it will perform in service. The base oil comes from two general sources: mineral or synthetic. The term *mineral* usually refers to base oils that have been refined from a crude oil source, whereas *synthetics* are usually the product of a chemical reaction of one or more selected starting materials. The finished fluids can also contain mixtures of one or more base oil types. Partial synthetic fluids contain mixtures of

mineral and synthetic base oils. Full synthetic fluids can also be mixtures of two or more synthetic base oils. As an example, mixtures of poly alpha-olefins and esters are commonly used in synthetic formulations today. Mixtures are generally used to tailor the properties of the finished fluid for a specific application or need. An overview of the general characteristics of different base fluids is shown in Table 1. Additional information regarding processing and composition of mineral and synthetic base fluids can be found in several sources (Mang and Dresel 2007; Pirro and Wessol 2001; Rudnik and Shubkin 1999). An overview of some of the general characteristics is described in the following section.

### Base Oil Characteristics

Mineral-based gear oils have been used successfully for many years in automotive and industrial gear drive systems. Mineral oil lubricants are petroleum-based fluids produced from crude oil through petroleum refining technology. Paraffinic mineral-based gear oils have viscosity indices that are commonly lower than most but not all synthetic-based gear oils. This usually means that the low-temperature properties of these mineral based lubricants will not be as good as for a comparable grade synthetic fluid. If low ambient temperatures are involved with the operation of the equipment, this should be factored into the decision process. At high temperatures, mineral-based lubricants are more prone to oxidation than synthetics due in part to the amount of residual polar and unsaturated compounds left in the base component. On the other hand, mineral-based lubricants will generally provide a higher viscosity under pressure than most synthetics



**Gear Lubricants, Fig. 1** Schematic of polymer character as a function of temperature

and will, therefore, produce thicker films. Probably the primary advantages of mineral-based oils over synthetic-based oils are their lower initial purchase cost and greater availability worldwide. Some of the inherent weaker properties of mineral oils, compared with a synthetic fluid, can be improved through the thickener and additive systems available today.

Synthetic oils differ from petroleum-based oils in that they are not found in nature, but are manufactured chemically and have special properties that enhance performance or accommodate severe operating conditions. Because they are manufactured, many of their properties can be tailored for specific needs through the choice of starting materials and reaction processing. Many synthetic oils are stable at high operating temperatures, have high viscosity indices (VI) (i.e., smaller viscosity changes with temperature variations), and low pour points. This means that equipment filled with most commercially available synthetic gear oils can be started without difficulty at lower ambient temperatures than those using mineral oils. Another key advantage is that they are inherently more stable at higher temperatures against oxidative degradation than their mineral counterparts, again owing this property to the uniformity and composition of the fluid structure.

Each type of synthetic lubricant has unique characteristics and all have limitations that should be understood. Such things as compatibility with other lubrication systems and mechanical components (seals, sealants, paints, clutches, etc.), behavior in the presence of moisture, lubricating qualities, and overall economics should be analyzed carefully for each type of synthetic lubricant under consideration for a given application. Synthetic lubricants, because of their viscosity-temperature characteristics and structure-influenced heat transfer properties, may improve gearbox efficiency and may operate cooler than mineral oils. Decreasing the operating temperature of

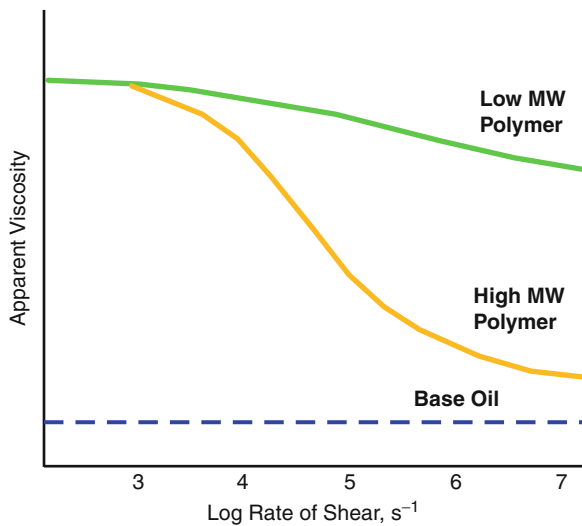
a gearbox lubricant is desirable. Lower lubricant temperatures increase the gear and bearing lives by increasing lubricant film thickness, and increase lubricant life by reducing the potential of oxidation.

### Thickeners

Thickeners, also known as viscosity modifiers (VM) or viscosity index improvers (VII), are quite common in automotive applications today, but are generally not used in industrial gear oil formulations. Thickeners are generally polymers that thicken the base oil to a much greater extent per unit volume of material than a conventional base stock, such as a bright stock or cylinder stock. At higher temperatures the molecule expands, creating a thickening effect. As the temperature decreases the polymer molecule tends to contract, minimizing the thickening effect. A schematic of this principle is shown in Fig. 1. The unique ability of these polymers to expand and contract as a function of temperature enables the finished blend to have much better viscosity-temperature characteristics.

Polymers are merely a chemical combination of one starting unit, known as a monomer, into many repeating units. The properties of the polymer are a function of the molecular weight (the number of repeating units) and the chemical structure of the monomer. Some of the more common polymer types used as viscosity modifiers include poly alpha-olefin, poly isobutylene, poly alkyl-acrylate and -methacrylate, and olefin copolymers, to name a few.

In addition to altering the viscosity-temperature properties of the finished fluid, the choice of polymer can also have an impact on the supporting film in the gear and bearing contact regions. The film formed in the contact will be a function of the temperature, pressure, and velocity of the mating surfaces. On the negative side, polymers are subject to mechanical and thermal shearing, which



**Gear Lubricants, Fig. 2** Effect of shear on polymer containing blends

results in a temporary and/or permanent loss of viscosity. The rate of loss is directly proportional to the molecular weight (MW) of the polymer, that is, higher MW polymers result in higher viscosity losses (see Fig. 2). Different polymer structures can also influence the response to pressure, temperature, and shear rate. Each of these parameters becomes important in the overall choice of a thickener.

### Chemical Properties – Performance Additives

Additives are typically a small (volume wise) but critical part of the overall formulation. Additive is a broad term that really encompasses many different chemicals, each providing performance or protection against certain types of damage or distress. These performance areas include, but are not limited to antiwear (AW), extreme pressure (EP) or antiscaff, ferrous corrosion, non-ferrous corrosion, demulsibility, oxidation, and foam inhibition. Chemicals impart or control these performance aspects in the application through reaction with the component surface or in the bulk oil phase. Most gear lubricants use a variety of chemicals in order to satisfy the many needs of the application. These chemicals must be selected properly, not only for the desired performance function, but for compatibility with the other chemicals in the package so that performance is not degraded through interactive effects.

Most commercial gear lubricants contain additives or chemicals that enable them to meet specific performance requirements. Typical additives include rust inhibitor,

oxidation inhibitor, antifoam, antiwear, and antiscaff agents. Many of the chemicals used to form the additive “package” are single function, but some can provide benefit in multiple areas. For example, certain phosphorus compounds, while primarily used for antiwear, can also provide protection against scuffing or function as oxidation inhibitors. Blends containing antiwear and antiscaff agents are generally referred to as EP oils.

Today, most gear lubricants are tailored to meet various industry-type specifications. These specifications generally dictate the composition of the chemical additive package by the performance level required. These specifications ensure the end user that a minimum level of performance can be expected from lubricants meeting their requirements. The specifications are usually designed to meet the needs of specific applications. For example, SAE J2360 is designed to cover rear axle hypoid applications while AGMA 9005-E02 and DIN 51517 cover a range of industrial applications (Surface Vehicle Standard SAE J2360 et al. 1998; ANSI/AGMA 9005-E02 et al. 2002; DIN 51517 (3 Parts) – Lubricants – Lubricating Oils (1 = Type C, 2 = Type CL, 3 = Type CLP) et al. 2002). Each has appropriate performance tests that must be met to qualify as an approved formulation. Regardless of whether the application is automotive or industrial, the functions and general chemistry used for each is similar enough to be classified in a broad category. It becomes the job of additive suppliers to balance and optimize the formulation for performance and cost with specific derivatives in each of the functional areas. Table 2 provides an overview of the broad chemical types that could be used to satisfy the stated functional need. Matching chemical structure to performance is a critical factor in the development of additives for commercial use.

### EHD Theory and Gear Lubricants

Over the years, significant contributions to the literature and our general knowledge about lubrication have been made, particularly with respect to the role of film thickness in lubrication (Dowson et al. 1970; Hamrock and Dowson 1977). Equations 1 and 2 represent the formulae derived by Dowson and colleagues for calculating the minimum film thickness for line and elliptical contacts, respectively. Combined with a knowledge of the mating surface combined roughness, one can then define a specific film thickness or lambda ratio (3) which in turn relates to the operating regime of the lubricant. This information provides some guidance about the type of lubricant one may need for an application.

$$h_{\min} = 2.65 * R * G^{0.54} * U^{0.70} * W^{-0.13} \quad (1)$$



**Gear Lubricants, Table 2** Generic chemical types used in gear lubricant formulations

Function	Chemical type	Common test method
Antiwear	Ashless and ash-containing phosphorus derivatives	ASTM D6121 (L37 test)
		FZG wear (various)
Antiscuff	Organo sulfur derivatives	ASTM STP512A (L42 test)
		ASTM D5182 (FZG scuffing)
Ferrous corrosion	Alkyl amines	ASTM STP512A (L33 test)
	Alkyl carboxylic acids	ASTM D665
Non-ferrous corrosion	S-N heterocycle derivatives, e.g., aromatic triazoles	ASTM D130
Oxidation	Phenolic	ASTM D5704 (L60-1)
	Aromatic substituted amines	CEC L-48-A-95 (DKA oxidation)
Antifoam	Poly siloxane derivatives	ASTM D892

$$h \min = 3.63 * R * G^{0.49} * U^{0.68} * W^{-0.073} * (1 - e^{-0.68k}) \quad (2)$$

where

$R$  = reduced radius of curvature ( $r_1 * r_2 / r_1 + r_2$ )

$G$  = dimensionless material parameter ( $\alpha * E$ )

$U$  = dimensionless speed parameter

( $\mu_0 * (u_1 + u_2) / 2ER$ )

$W$  = dimensionless load parameter ( $w/ERL$ )

$$\lambda = h \min / \sigma \quad (3)$$

where

$\sigma$  = composite roughness

The lubricant plays a direct role in these equations through the material and speed parameters. The material parameter uses the pressure-viscosity coefficient ( $\alpha$ ) of the lubricant while the speed parameter takes into account the effective viscosity ( $\mu_0$ ) at the point of entrainment into the contact. However, these equations were developed empirically in the 1960s and 1970s and as such do not really account for shear losses associated with many modern polymer-containing formulations. More recent work by Bair and Khonsari has addressed the effect of shear-thinning fluids and developed correction factors to enhance the estimates (Bair 2005; Kumar and Khonsari 2008).

Another factor not accounted for with these equations is the influence from the chemical additive. The foregoing equations address the physical contributions of the lubricant to the overall performance. To date, there is no simple way to address the chemical impact of

a formulation on performance. Even with advanced analytical tools available today that can provide detailed structure composition for the chemicals used in formulations, the performance outcome is driven in many cases by subtleties involving the hardware and operating conditions as well as the chemical structure. Additionally, interactions occur within the oil and at the surfaces so the chemical entity becomes a dynamic or transient species in the process. Thus, the only true method to determine specific performance is through testing. The test methods described in the SAE J2360, AGMA 9005-E02, and DIN 51517 standards and others usually provide a good estimate of expected performance from the lubricant physical and chemical properties, as might be encountered in service. Equipment manufacturers will many times use these standards as a starting point and then add specific requirements to address the needs of their particular equipment.

## Key Applications

While gear lubricants have found uses in many varied applications, the primary areas of use for these materials are in vehicle drive axles and manual transmissions and in enclosed drives used with stationary equipment.

## Cross-References

- [Gear Lubricants](#)
- [Gear Lubrication](#)
- [Rheology – Viscosity Index](#)
- [Rolling Bearing Lubricants](#)
- [Shear Dependence of Viscosity](#)
- [Transmission Lubricants](#)
- [Viscosity Index Additives](#)

## References

- ANSI/AGMA 9005-E02, American National Standard, Industrial Gear Lubrication, American Gear Manufacturers Association, Dec 2002
- S. Bair, Shear thinning correction for rolling/sliding elastohydrodynamic film thickness. *Proc. Inst. Mech. Eng.: J. Eng. Tribol.* **219**(1), 69–74 (2005)
- DIN 51517 (3 Parts) – Lubricants – Lubricating Oils (1 = Type C; 2 = Type CL; 3 = Type CLP), Deutsches Institut fuer Normung, e.V., 2002
- D. Dowson, Elastohydrodynamic lubrication, in *Interdisciplinary Approach to the Lubrication of Concentrated Contacts*, Special Publication No. NASA-SP-237, National Aeronautics and Space Administration, Washington DC, 1970, p27 ff
- B. Hamrock, D. Dowson, Isothermal elastohydrodynamic lubrication of point contacts – Part III – fully flooded results. *Trans. ASME, J. Lubr. Technol.* **99**(2), 264–276 (1977)
- P. Kumar, M.M. Khonsari, EHL circular contact film thickness correction factor for shear-thinning fluids. *J. Tribol.* **130**(4), 041506 (2008)
- Th Mang, W. Dresel, *Lubricants and Lubrication*, 2nd edn. (Wiley-VCH, Darmstadt, 2007)
- D.M. Pirro, A.A. Wessol, *Lubrication Fundamentals*, 2nd edn. (Marcel Dekker, New York, 2001)
- L.R. Rudnik, R.L. Shubkin (eds.), *Synthetic Lubricants and High Performance Functional Fluids* (Marcel Dekker, New York, 1999)
- Surface Vehicle Standard SAE J2360, Lubricating Oil, Gear Multipurpose (Metric) Military Use, Society of Automotive Engineers, Nov 1998

## Gear Lubrication

DENNIS A. LAUER

Klüber Lubrication North America L.P., The Villages,  
FL, USA

## Synonyms

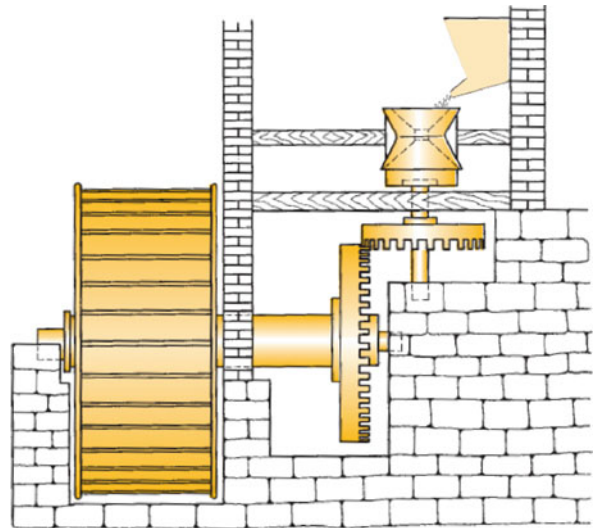
Gearing energy losses; Gear friction reduction; PAG gear oils; Polyglycol gear oils; Synthetic gear oils; Wormgear lubrication

## Definition

Gear wheels and simple gear units were common, even in ancient times, and were frequently seen in grain mills (Fig. 1), water pumping stations, and wind mills.

Even today, many centuries after the gear wheel was invented, there is nothing better, more compact or efficient than a gear unit when it comes to transferring power, converting torque, and adapting speed. It is, therefore, not surprising to find gears in all fields of technology, including turbo-generator drives with an output of 100 MW (134,000 hp) and more, ball mill and kiln drives, or small and miniature low-output drives used in mechanical equipment and computers.

Like all machine elements under repeated stress, gears have a predictable life, with fatigue being the primary



**Gear Lubrication, Fig.1** Roman grain mill with a right-angle gear (Figures courtesy of Klüber Lubrication München KG)

determining factor. Modern gear design's objective is to achieve the most cost-effective service life possible under a given set of operating conditions. It is not uncommon for properly designed, lubricated, and maintained gear drives to exceed 20 years of continuous service life. To achieve optimum service life, the gear designer must take into account the forces to be transmitted, the effects of shock loads, and operating conditions including speed, temperature, and environment. With this information, the designer can select the proper metallurgy, gear type, pitch diameter, tooth form, size, modifications, lubrication requirements, and, ultimately, the machining and surface finishing operations used in the manufacturing process. The lubrication method, lubricant type, amount, and frequency of application will vary depending on gear drive type, for example, open, enclosed or semi-enclosed, and operating conditions and environment.

Increased efficiency is mainly achieved by means of gear materials that are more resistant to wear (case hardened, alloy steel), optimized tooth geometry, improved flank machining and finishing methods, and lubrication. The power-weight ratio of gear units can also be improved by means of torque division in planetary gears, lightweight construction such as light metal or plastic casings and hollow shafts, or by increasing the thermal limit that often lies below their mechanical load limit by using synthetic lubricating fluids.

Lubrication requirements, in terms of antiwear and anticuff properties as well as high temperature resistance, increase when the power/weight ratio of the gear

unit increases. Excess heat must be dissipated from gears with a very small housing surface. This often results in higher operating temperatures, which has a negative effect on the service life of gears and lubricants.

Measures to decrease power losses and reduce temperatures, such as the reduction of friction losses in gear components as well as gear cooling, continue to gain importance as the cost of energy and environmental concerns increase. Premium high-performance lubricants, mineral oil or synthetic based, with a low traction coefficient and high resistance to increased temperatures and oxidation, can often be a cost-effective way of decreasing power losses and reducing operating temperatures. In many cases, complicated cooling installations are no longer necessary.

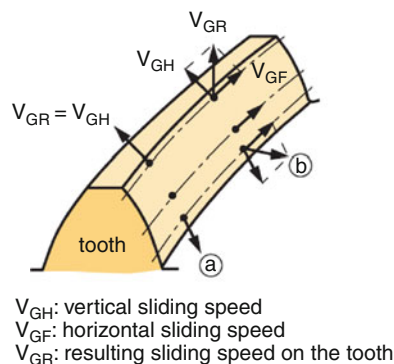
When the gear designer is successful, the gear design does the job it is meant to do with some margin of safety. The lubrication engineer succeeds when the lubricant and lubrication method prevents lubrication-related failures. Other forms of failure result from flaws in the design that may be based on metallurgy, geometry, misalignment, insufficient tooth strength, or surface durability.

## Scientific Fundamentals

### Gear Lubrication Basics

#### Types of Movement and Speed

With all types of gears, the tooth flanks perform rolling and sliding movements while the teeth are meshing. The sliding percentage of the combined rolling/sliding movements varies depending on the type of gear. Sliding/rolling speed ratio is of vital importance for the load to which the lubricant is subject.



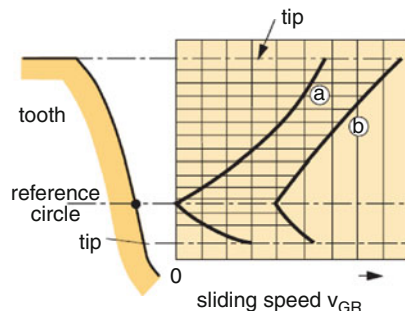
In gears where the rolling movement is predominant, such as spur gears and helical gears, the tooth flanks and the lubricant are subject to less sliding than in gears with a predominant sliding movement such as worm gears and hypoid gears. Figure 2 illustrates the speed conditions in spur gears and helical gears. Point (a) shows that in spur gears the sliding vector is only vertical. It is zero at the pitch line, pure rolling friction, and increases in the direction of the tip of the tooth. Point (b) illustrates that in helical gears there is also a sliding (spiral) movement in the horizontal direction and that there is even a certain percentage of sliding movement at the pitch line.

This shows why the ratio between the sliding and the rolling speed is relatively high in worm gears and hypoid gears with offset shafts. The resulting sliding vector direction is not constant but changes with vertical position. The “wiping” movement considerably impairs the formation of a lubricant film under pressure. Load on the tooth flanks and the lubricant increases as the sliding percentage grows.

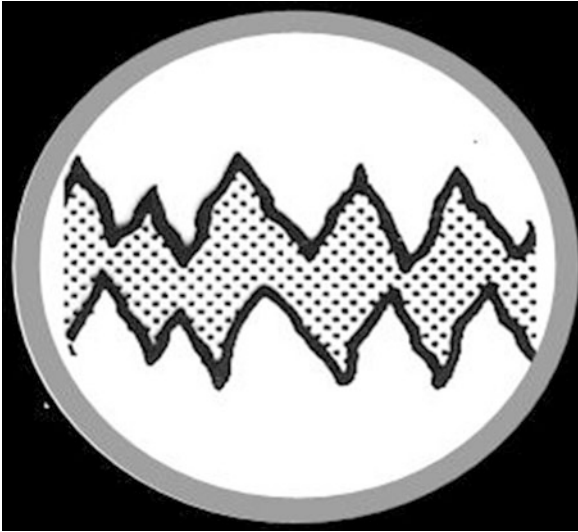
### Lubrication Condition

#### Full Fluid Film Lubrication

Full fluid film lubrication would be the ideal condition of lubrication (Fig. 3). The meshing tooth flanks would be completely separated by a lubricant film. Due to the gap geometry and the various movements, however, full fluid film lubrication on the meshing flanks is only achieved on sufficiently large flank areas under certain operating conditions; mixed friction prevails on the other areas. The full fluid film area size depends mainly on the tooth flank geometry and is the largest in spur and bevel gears.



**Gear Lubrication, Fig. 2** Sliding speeds and vectors on tooth flanks (Figures courtesy of Klüber Lubrication München KG)



**Gear Lubrication, Fig. 3** Full fluid film (Figures courtesy of Klüber Lubrication München KG)

The formation of a load-carrying and separating lubricant film around the pitch line can be explained by means of the elastohydrodynamic lubrication (EHL) theory (Fig. 4), which is based on the following:

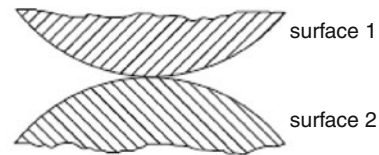
- High pressures up to 1,500 MPa (218,000 psi) occur in the contact zones between the tooth flanks. They increase oil viscosity dramatically.
- The contacting surfaces deform elastically.

High surface pressure between the tooth flanks results in a sudden increase in the oil film's viscosity, making the lubricant so viscous that it cannot readily escape. The oil film pressure deforms the tooth flanks elastically in the contact points, and flattens and enlarges the contact area, which leads to a greater lubricant gap and a thicker oil film. Optimum conditions for a separating film are present in spur gears, due the high percentage of rolling movement and the line contact when the teeth mesh. Film thickness increases as the pressure-related viscosity and the rolling speed rise. Shearing of the lubricant in the inlet region of the contact, however, heats the lubricant and reduces the viscosity somewhat. The effect is usually significant only for high-speed gears.

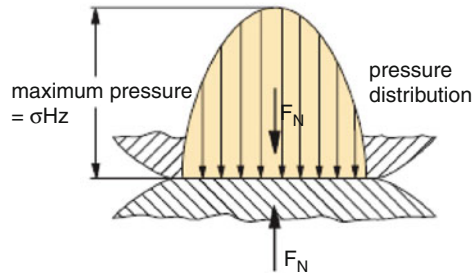
#### Mixed Lubrication

Mixed lubrication is the lubrication condition that includes both fluid tractional losses due to shearing of the oil film and frictional forces at surface asperities. Surfaces of the meshing teeth are not completely separated by a fluid film and are in direct contact in some areas.

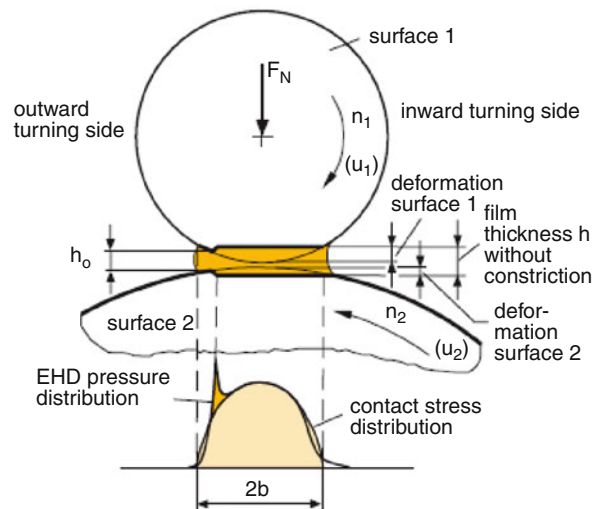
#### a no load



#### b load, no movement



#### c load and rotation with lubricant film

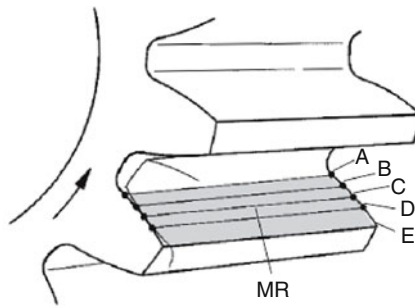
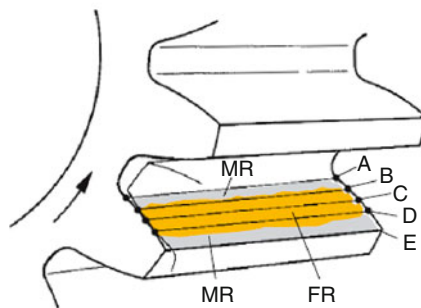


**Gear Lubrication, Fig. 4** Formation of a lubrication gap in elastohydrodynamic lubrication (EHL) (Figures courtesy of Klüber Lubrication München KG)

#### Mixed Friction

Mixed friction occurs in cylindrical and bevel gears operating at low peripheral speeds and subject to high specific load (Fig. 5). At increased peripheral speeds, mixed friction is replaced by fluid friction (Fig. 6). Worm and hypoid gears usually operate under mixed friction conditions.

A formulated lubricating oil with antiwear or antiscaff additives forms reaction layers (boundary layers) on the flank surfaces. Antiscaff additives are intended to

**a** low peripheral speed**b** high peripheral speed

A ... E: meshing points  
in acc. with DIN 3960

A initial meshing point

B internal individual meshing  
point of the driving gear,  
external individual meshing  
point of the driven gear,

C pitch point

D external individual meshing  
point of the driving wheel,  
internal individual meshing  
point of the driven wheel,

E final meshing point

FR full fluid film lubrication

MR mixed lubrication

**Gear Lubrication, Fig. 5** Mixed friction (Figures courtesy of Klüber Lubrication München KG)

prevent or at least limit metal-to-metal contact and reduce the potential of scuffing, and severe wear. Sulfur/phosphorous-type antiscuff additives are the most common for extreme pressure. Chemical reactions with the tooth material produces iron sulfide and iron phosphate layers. Pressure, temperature, type, and quantity of the additives impact the course and intensity of the tribo-chemical reactions. The additives are usually combined in such a way that they become active with an increasing tooth flank temperature. As compared to the base materials, the reaction layers have low shear strength that ensures a continuous process of wear and renewal of the boundary layers.

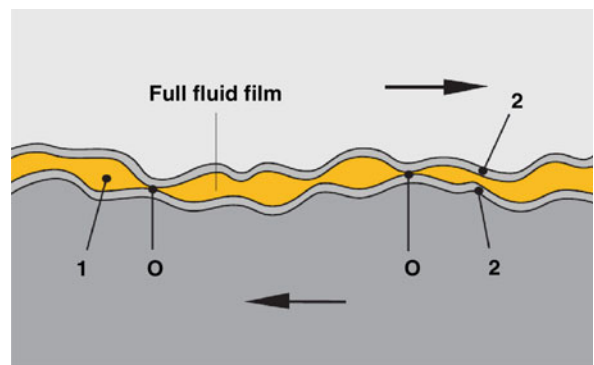
## Key Applications

### Types and Methods of Lubrication

#### Oil Lubrication

##### Splash Lubrication

The easiest way to ensure continuous lubrication is a lubricant sump in the gear housing into which the teeth are immersed (Fig. 7). It is therefore the most



O Surface layer friction

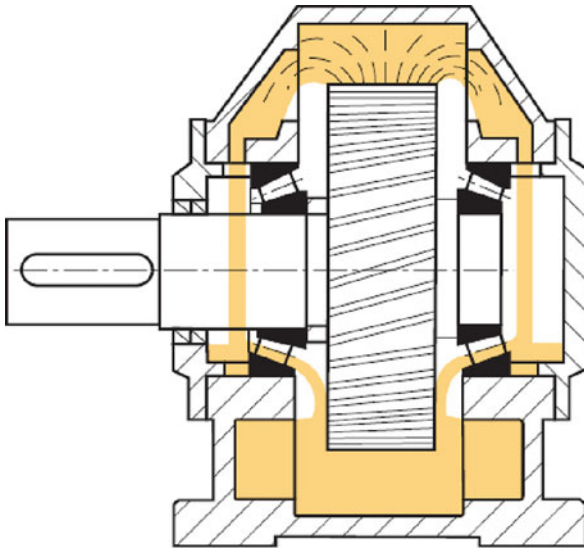
1 Lubricant

2 Surface reaction layer

**Gear Lubrication, Fig. 6** Typical friction areas on tooth flanks (Figures courtesy of Klüber Lubrication München KG)

common lubrication method for closed, oil-tight gear systems. It is especially economic, simple in design, and reliable. A lubricant sump's cooling effect is sufficient for a wide range of applications and can be improved by





**Gear Lubrication, Fig. 7** Splash lubrication. Oil lubrication of rolling bearings via intermediate reservoirs (Figures courtesy of Klüber Lubrication München KG)

means of special housing designs, for instance, cooling ribs, or auxiliary units such as cooling fans or water cooling lines.

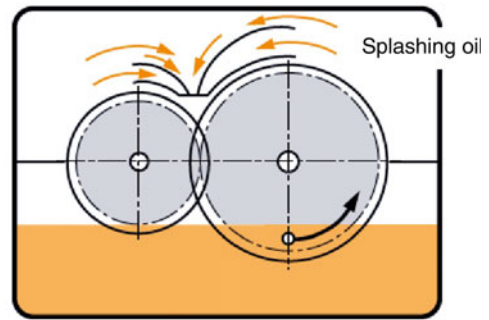
One wheel per gear stage usually plunges into the oil sump and carries some oil to the meshing zone in splash lubrication. Oil thrown off the gears and shafts, due to centrifugal forces, either runs into special oil return lines and subsequently is carried to the bearings that also require internal lubrication, or runs directly back into the sump along the housing walls. Splashing oil to a certain extent also lubricates the gears that are not immersed. In some cases, the shafts are equipped with additional oil slingers to intensify the splash effect.

Oil splash lubrication without any additional design measures is feasible up to a peripheral speed of approximately 20 m/s (3,937 fpm). Oil guides, including baffle plates and oil pockets, are required for increased peripheral speeds (Fig. 8). Splash lubrication with additional design measures is efficient up to a peripheral speed of 60 m/s (11,811 fpm) in tests with single-stage spur gears.

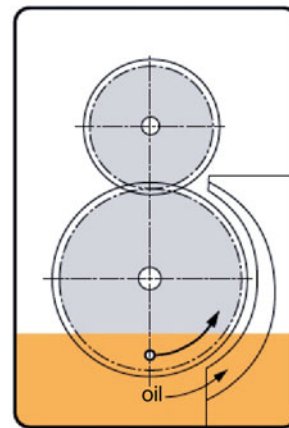
#### Oil Level and Immersion Depth

It is important with splash lubrication that a certain oil level is maintained at all times in order to prevent damage and ensure reliable operation. If the oil level is too low, it may result in starved lubrication, inadequate heat dissipation, increased wear, foam, and noise. Churning

Horizontal arrangement



Vertical arrangement



**Gear Lubrication, Fig. 8** Examples of oil guide plates (Figures courtesy of Klüber Lubrication München KG)

losses may increase, resulting in a higher operating temperature and a decrease in overall drive efficiency if the oil level is too high. Increased operating temperature accelerates oil aging and decreases the oil's service life. Noise and the generation of foam may also occur if an oil level is too high.

When the peripheral speed is increased, the depth of immersion is reduced in order to keep churning losses to a minimum. With higher peripheral speeds it becomes more difficult to wet the tooth flanks sufficiently, and the oil level is reduced due to oil splashing. In the case of very high peripheral speeds it is generally best to use a circulating oil system. A great difference in wheel diameters in the individual gear stages may also be a reason for deviating from the immersion depths listed in Table 1.

#### Combined Splash and Circulation Lubrication

If additional cooling ribs, fans, or ducts in the housing are not enough to dissipate the heat from the power



**Gear Lubrication, Table 1** Recommended depth of immersion

Gear type	Speed or worm position	Immersion depth
Spur and helical gears	Peripheral speed up to 5 m/s (984 fpm)	3–5 times the module
	Peripheral speed > 5 to ~ 20 m/s (>984 to ~ 3,937 fpm)	1–3 times the module
Bevel gears		Immersion of teeth over the entire width of the wheel
Worm gears	Worm on the top	Wheel immersed to approx. 1/3 of its diameter
	Worm on the bottom	Wheel immersed to approx. the center of the meshing zone
	Worm at the side	1/2 Wheel immersed to at least worm height

losses in a splash-lubricated gear, heat can also be dissipated by using a combined splash and circulation lubrication system. To dissipate the heat, the easiest method is to install a pump driven by the gear that takes a specific oil quantity out of the oil sump and returns it to the gear via an oil cooler. Electrically operated oil pumps are also used to disperse the heat. Thermostatically controlled coolers and filters installed in the oil circuit make it possible to extend the oil change intervals. The depth of wheel immersion is determined as described for splash lubrication. The oil fill quantity has to be higher because a certain amount of oil is always circulating.

#### **Pressure Circulation Lubrication (Injection Lubrication)**

When the peripheral speed of the gearing is too high for splash lubrication or when the drive is equipped with plain bearings, pressure circulation lubrication is used. This lubrication method is designed into gear drives used in critical path and reliability-critical applications as well as drives designed to achieve an extended service life between repairs. Pressure circulation lubrication is suitable for even the highest peripheral speeds encountered in gear systems, approximately 250 m/s (49,200 fpm). Oil is sprayed onto the tooth flanks via slotted or drilled nozzles. In very high-speed applications, such as very large, 20 MW (26,800 hp), integral gear-driven centrifugal compressor or helicopter transmissions, the nozzles are designed to produce a high-pressure small diameter stream of oil that can pass undisturbed through the turbulent atmosphere at the mesh area of the gears. This is commonly referred to as oil jet or oil injection lubrication. Whether applied by a low-pressure fan or stream-shaped spray or by a higher pressure jet stream, the oil is directed into the contact zone, either at the inlet or exit to the meshing zone, or both. It is assumed that oil directed into the inlet of the

meshing zone is more beneficial to the lubrication process, while oil directed into the exit of the meshing zone increases the cooling effect.

The quantity of oil directed into the gear contact zone or zones depends on the amount of heat to be dissipated. As a rule of thumb, 0.5–1.0 l/m per cm (0.33–0.66 gpm per in.) of tooth width is recommended. Determining oil circulation quantity is the combination of the lubricant volume required for the gear wheels, plus the quantity required for the bearings. Depending on the type of circulation, a distinction is made between wet and dry sump lubrication. In wet sump lubrication, the oil reservoir is the oil sump in the housing from which the oil is brought to the friction points. In dry sump lubrication, the oil that returns from the friction points is collected in the housing and then transferred into a separate oil container, from where it is brought back to the friction points. Dry sump lubrication is used if the oil volume required is too much to be accommodated in the gear housing or when maximum gear drive efficiency is desired.

#### **Grease Lubrication**

Grease lubrication is normally used for gear drives that are not oil tight, for safety reasons, leakage, or for oil tight gears installed in a position limiting or preventing maintenance and repair. Gear drives suitable for lifetime lubrication are suitable for grease lubrication. As grease is limited greatly in the amount of heat it will or can dissipate as compared with oil, grease lubrication is typically restricted to low and medium capacity gears operating in short or intermittent intervals. Gear motors and all types of small and miniature gears, such as power tools, household and office equipment, and automotive servo drives are the primary applications for grease. However, there are instances where specialty high performance semi-fluid greases, NLGI 00 and 000, have been successfully used for extended service periods in

**Gear Lubrication, Table 2** Lubrication method based on gear speed

Peripheral speeds, m/s (fpm)	Application Method
Up to 2 (394)	Manual application
>2 and up to approx. 8 (>394 to approx. 1,575)	Splash or spray lubrication
> 8 to approx. 11 (>1,575 to approx. 2,165)	Spray lubrication

heavy-duty and severe-duty gear drive applications including gantry cranes in steel mills and tandem mill drives in cane sugar processing plants.

The common method of application is splash lubrication for grease-lubricated gears. Gear drives with a low specific load that are mainly used to transfer movements over a limited period of time, or gears designed for a limited operating period are usually lifetime lubricated. With this application method, the peripheral speed is limited to 2 m/s, 394 fpm (Table 2).

Splash lubrication with greases is generally restricted to low- to medium-capacity gears. The exceptions have typically been limited to heavy-duty rotary kiln and grinding mill open gear drives, and enclosed gear drives using straight immersion (splash) or idler immersion (open gear drives only) lubrication systems. In these instances, semi-fluid, high-performance, specialty greases containing solid film lubricant additives with or without additional antiwear and antiscaff additives are used. Peripheral speeds range from approximately 1.2 to 5.0 m/s (236 to 984 fpm). A thorough technical review of the application involving the gear drive user, the OEM, and lubricant provider is essential before using grease to splash lubricate heavy-duty enclosed or open gear drives.

Since the development and dissipation of heat largely depends on the individual operating conditions, such as continuous or intermittent operation, the case design, and material(s) of construction, it is not possible to indicate a specific capacity limit for the grease lubrication of enclosed gear drives. Unless the gears have a very low capacity, it is recommended to carry out tests to establish the gear heating rate. Gears with high peripheral speeds operating continuously and/or used to transfer power require a grease with a high base oil content to ensure heat dissipation; a semi-fluid grease of consistency NLGI grade 000 or 00, or as a maximum a very soft grease of NLGI grade 0. For continuous operation the rule of thumb peripheral speed limits are provided in Table 3.

**Gear Lubrication, Table 3** Correlation between the Peripheral Speed and NLGI Grade

NLGI grade	Peripheral speed, m/s (fpm)
000	6–8 (1,181–1,575)
00	4–5 (787–984)
0	2–3 (394–591)

Semi-fluid greases have a less favorable flow behavior than oils. Excess grease thrown off the gear returns to the lubricant sump at a much slower speed. It is therefore important to fill the grease immersion bath to a higher level than for oil lubrication in order to prevent starved lubrication. Depending on the grease's consistency, the approximate depth of immersion should be between 1 and 3 times the tooth depth. Splash grease lubrication of an enclosed gear drive at peripheral speeds above 8–10 m/s (1,575–1,969 fpm) in continuous operation is possible if the following steps are taken:

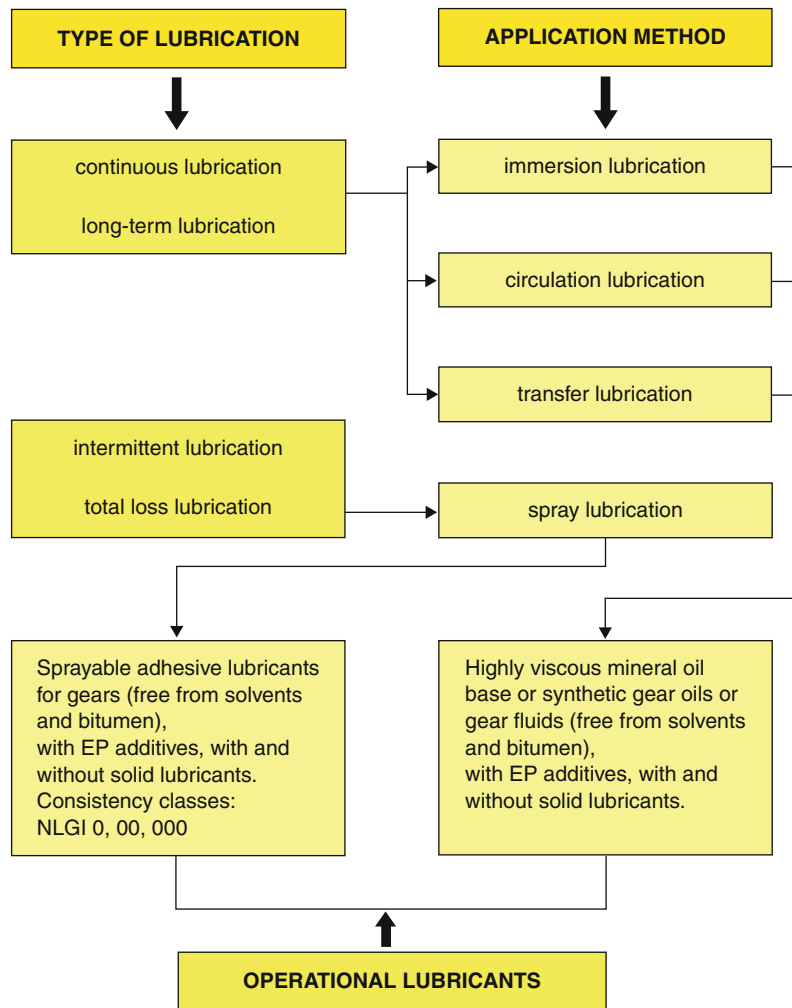
- Use a gear grease of NLGI consistency grade 00 or 000 and fill the case to approximately 30–40% of its volume.
- Use a grease of NLGI consistency 0, 1, or 2 and fill the case almost completely.

Low-capacity gears in short-term or intermittent operation can be lubricated with grease up to a peripheral speed of 25 m/s (4,920 fpm) if the second method is applied. Regardless of the type of gear drive to be splash lubricated, only use a grease formulated with a base oil, thickener, and additives appropriate for gear lubrication; correct ISO viscosity grade for the speed and load involved, highly shear stable over its anticipated service life, and the correct level of antiwear and/or antiscaff protection.

#### Lubrication with Adhesive Lubricants

Adhesive lubricants are mainly used for the lubrication of open gears and gear systems of large to very large dimensions, such as drives of rotary kilns, ball mills, lifting cylinders, cranes, and construction machines. Depending on the operating mode and/or conditions and if the drive is designed with/without a cover or with/without lubricant reservoir, the lubricant is applied:

- Once (or several times at very long intervals)
- Continuously or near continuously (long term lubrication)
- Intermittently (total loss lubrication)



**Gear Lubrication, Fig. 9** Lubrication types and application methods common with lubricants for large open gears (Figures courtesy of Klüber Lubrication München KG)

Figure 9 shows the correlation between the types of lubrication and the application methods. The type of lubrication and the suitable application method is determined by the lubricant used. Adhesive lubricants used on open gears can be fluid or semi-fluid and therefore have an impact on the application method. Asphaltic or residual oils are of three types: straight (without compounding), compounded, and solvent cutback. All of the oils in this group are black, viscous, heavy, and sticky; and they leave a very tenacious film. Newer technology open-gear lubricants include sprayable greases and highly viscous fluids. They are widely used on open gears that operate at peripheral speeds not exceeding 11 m/s (2,165 fpm). Even when they are enclosed by a housing, many large, slow-speed drives are lubricated as “open gears.”

## Cross-References

- [Friction in Gears](#)
- [Gear Efficiency](#)
- [Gear Lubricants](#)
- [Wear in Gears](#)

## References

- H.P. Bloch, J.B. Williams, Synthetic lubricants measure up to claims. *Chem. Eng.* **102**(1), 127–130 (1995)
- Die Schmierung von Zahnradgetrieben, 9.3/9.4 d, Ausgabe 03, 1995, Klüber Lubrication München KG.
- P.S. Korosec, *Gear Oils, Synthetic Lubricants and High Performance Functional Fluids*, 1st edn.
- D.A. Lauer, Gear oil classification and specification. *Gear Technol.* May–June 1995
- Lubrication, Vol 66, No 2 & 3, (Texaco, 1980)

- Lubrication of Gear Systems*, 9.3/9.4e, edn 11.03 (Klüber Lubrication München KG)
- U. Mann, *Synthetic oils for worm gear lubrication*. AGMA, Technical Paper 99FTM17, 1999
- K. Michaelis, B.-R. Höhn, *Influence of lubricants on power loss of cylindrical gears*. STLE, Preprint No. 93-AM-4B-1, 1993
- Personal communication with Kline and Company, Bill Downey, February 1997
- Plant Engineering Magazine's Exclusive Chart of Synthetic Lubricants*, Plant Engineering, Cahners Publishing Company, Reed Elsevier, Inc.
- Product Information S-002-01, SEW Eurodrive, April 1990
- L.R. Rudnick, *Synthetics, Mineral Oils, and Bio-Based Lubricants, Chemistry and Technology* (CRC Press, Boca Raton, 2006)
- Transmission Engineering, Selection and Application of Gear Lubricants* (Klüber Lubrication München), 1995
- H. Winter, P. Oster, Influence of lubrication on pitting and micro-pitting resistance of gears. *Gear Technol.* March/April 1990

## Gear Manufacturing

CHARLES D. SCHULTZ

Beyta Gear Service, Winfield, IL, USA

### Synonyms

[Gear cutting](#); [Gear grinding](#); [Gear molding](#); [Gear shaping](#); [Gear stamping](#)

### Definition

Gear manufacturing refers to the process by which raw material is converted into a useable gear.

### Scientific Fundamentals

It is not clear when gear manufacturing first separated itself from the more general mechanical arts. Once specialized tooth cutting machinery began to appear in the latter half of the nineteenth century, it became more difficult for the general foundry, blacksmith, or machine builder to competitively produce gears. Modern gear manufacturers have become further specialized by the types of gears they make and the markets they serve. Low-volume, high-precision gears require a completely different factory than high-volume, commercial gears. Material selection and heat treatment method are the primary determinant of the process used. Molded plastic or sintered metal gears are most commonly used in high-volume, low-load applications. Surface-hardened and ground-steel gears predominate in the low-volume, high-load applications. A variety of approaches are employed in the middle ground of these extremes.

Figure 1 is a flowchart comparing the manufacturing process used to make cut-steel gears.

### Key Applications

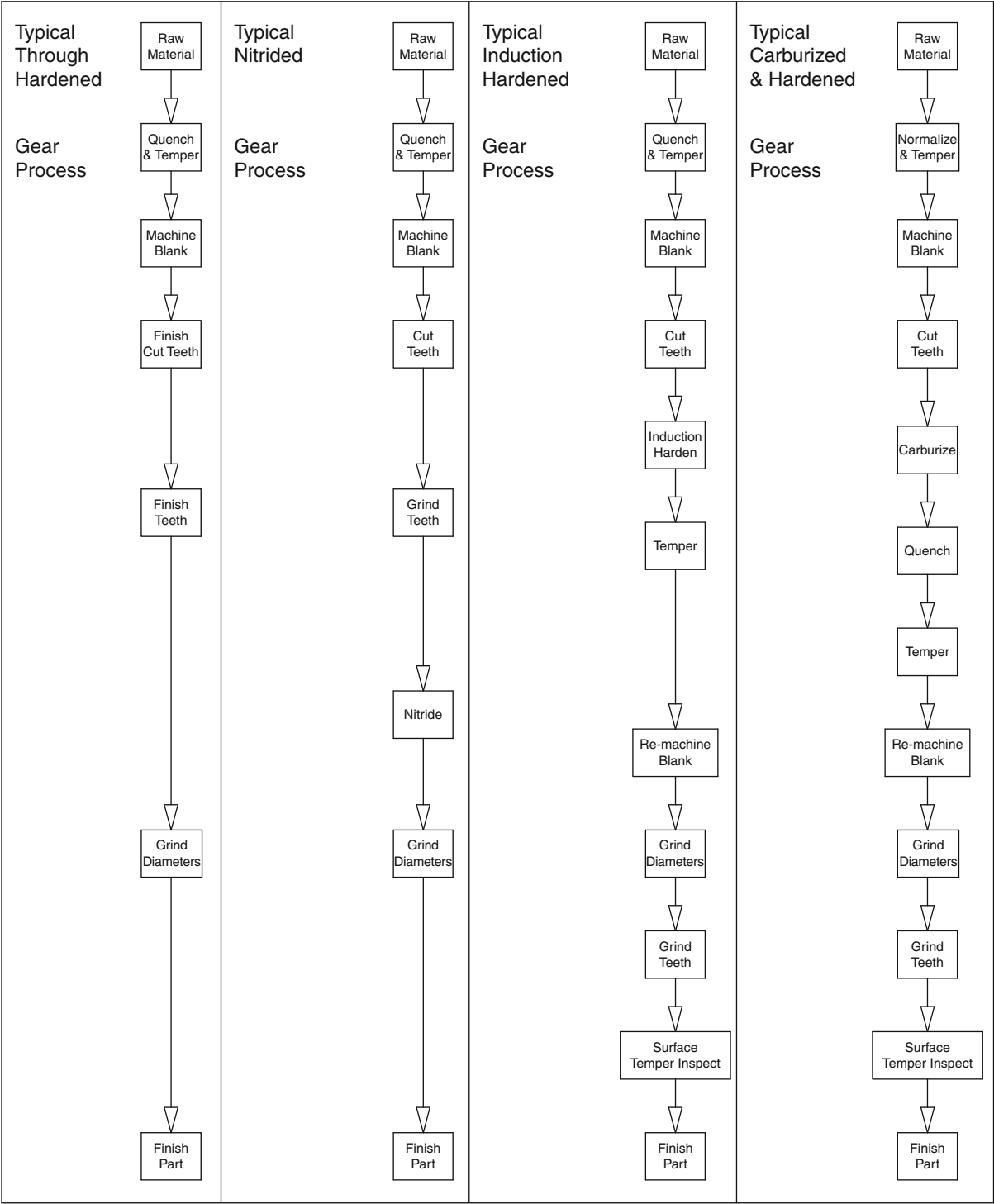
Cut gears typically start with raw material in a forged or rolled shape. After the raw material is cut into a size suitable for making the desired part, it may undergo a preliminary thermal processing to insure a uniform hardness and micro-structure for machining. Non-carburized parts are heated above the critical temperature (above 1,650 °F), quenched in oil or polymer, and then re-heated (tempered) to an intermediate temperature (above 1,150 °F) and slow cooled (50–100 °F per hour) to below 600 °F to establish the desired through hardness. Carburized parts receive a similar treatment with the objective of reducing the part hardness below 220 HB for ease of machining and stability during later heat treat operations.

Blank preparation is also quite similar, regardless of final heat treat condition, except for allowances for heat treat distortion and growth. Induction-hardened and carburized parts experience varying degrees of distortion during post gear cutting heat treat operations, and blank sizes must be adjusted to provide sufficient material to remove that distortion. It is not uncommon for some dimensions to grow as much as .001" per inch during heat treat while others shrink by a like amount. Inconsistency between part batches remains a difficult problem for gear manufacturers, so stock allowance practices are often considered proprietary information. Different alloys exhibit different growth and distortion characteristics and some manufacturers may have a preference for one alloy over another as a result of their different experiences.

Stock allowance also becomes a consideration at the tooth cutting operation. For parts that will have the teeth finished by shaving, honing, lapping, or grinding, the manufacturer must decide how much stock will be needed for that operation and what effect heat treat distortion and growth will have. Again, this information is valuable to a manufacturer and may be considered proprietary. Some will also make adjustments to the shape of the gear teeth at this step to reduce the amount of correction needed later. Adding a crown, end relief, or taper to the teeth are common "green" adjustments. Bevel gears, which are seldom finished after heat treat, are especially likely to have adjustments made at the tooth-cutting operation.

Through-hardened and nitrided gears usually have their teeth finished right after cutting. Through-hardened gears don't require further heat treating and as a result have no risk of growth or distortion. Nitrided gears do not

Gear Manufacturing Flow Chart



Gear Manufacturing, Fig. 1 Gear manufacturing flow chart

get quenched and are at low risk of distortion or growth. In addition, nitrided case depths are very thin and grinding after nitriding is likely to remove much of the hard layer.

Despite the concerted efforts of many generations of gear engineers, distortion and growth during induction hardening and carburizing and hardening operations remain a processing problem. As a result, these parts usually require re-machining after heat treat to re-establish locating surfaces for tooth grinding. Some surfaces of the parts may have been protected from hardening by the application of plating, paint, or mechanical masking prior to heat treat. Care must be taken during re-machining operations to insure equal stock removal in the tooth grinding process. Re-establishing the locating surfaces with respect to the gear pitch diameter can be a time-consuming test of operator skill.

Tooth grinding is a critical step in making high-quality gears. Even the best machinery, tooling, and operators can damage the tooth surfaces by cracking and/or overheating the tooth surfaces. Surface temper inspection and crack checking are an integral part of gear grinding practice. Failure to detect these problems early in the process can result in expensive scrap and delayed shipments. The warranty cost of cracked or burned gears getting into service can far exceed the expenses associated with proper inspection and process adjustment.

Depending upon drawing requirements, parts may undergo further processing after being essentially “finished” dimensionally. Some additional operations are polishing, super finishing, shot peening, coating, balancing, and size matching.

## Non-Cut Gear Manufacturing Processes

High-volume gears are usually made by processes other than cutting. Plastic molding, powder metal sintering, and stamping are popular methods for making gears by the thousands. Each process has unique advantages and disadvantages that must be considered during design, planning, and manufacturing.

## Cross-References

- [Gear Cutting Tools](#)
- [Gear Manufacturing Machines](#)
- [Gear Surface Treatment](#)

## References

- D.W. Dudley, *Gear Handbook*, McGraw-Hill Library of Congress Catalog Card Number: 61-7304
- D.W. Dudley, *Practical Gear Design*, McGraw-Hill Library of Congress Catalog Card Number: 53-11476

- D.W. Dudley, *The Evolution of the Gear Art*, American Gear Manufacturers Association Library of Congress Catalog Card Number: 72-78509
- C.D. Schultz, *An Introduction to Gear Design*, Privately published 1987–1997. Available for free download at [www.beytagear.com](http://www.beytagear.com)

## Gear Manufacturing Machines

CHARLES D. SCHULTZ

Beyta Gear Service, Winfield, IL, USA

## Synonyms

[Electric discharge machining \(EDM\) for gear manufacturing](#); [Electrical discharge machining \(EDM\)](#); [Grinding for gear manufacturing](#); [Hobbing](#); [Milling](#); [Planing](#); [Shaping](#); [Shaving](#), [broaching](#); [Water jet cutting](#)

## Definition

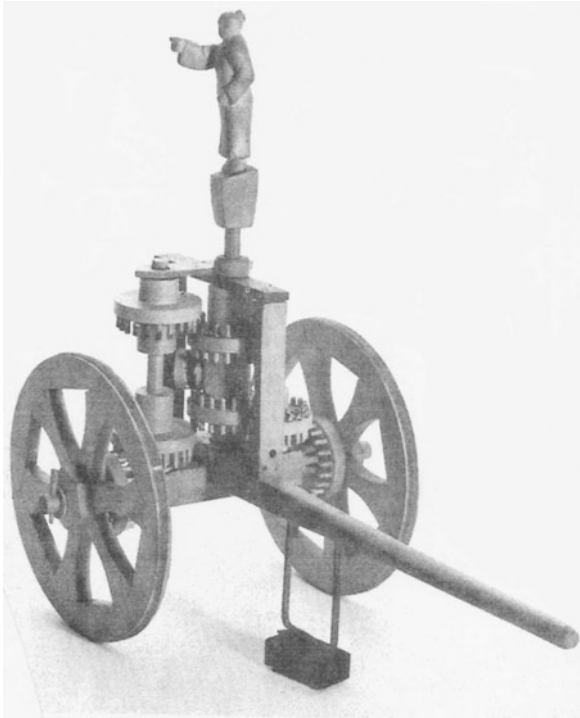
Gear manufacturing machines are used to make gear tooth surfaces. A wide variety of methods has been and continues to be employed to make gear teeth. This article will provide a brief history of this 5,000-year-old art, review the types of machines currently used to make gear teeth, and highlight recent advances in the technology. References are provided for those seeking more detailed information.

## Scientific Fundamentals

It is difficult to state with certainty when gears were first “invented.” Dudley notes that Aristotle wrote about gears (330 BCE) as if they were commonplace and that the relative complexity of the Chinese “South Pointing Chariot” (Fig. 1), circa 2600 BCE according to some accounts, indicates far earlier origins. Archaeological evidence of the use of gears in religious artifacts and more utilitarian applications such as grist mills and water lifting equipment is widespread. The writings of many of antiquity’s great thinkers include information on geared mechanisms. By 100 BCE, gears were being made in wood and metal. Spur gears, racks and pinions, worm gears, and primitive right-angle “bevel” gears were already in use.

The first “gears” were wooden pegs fitted into the rims of wheels. As the “South Pointing Chariot” shows, very complex motions can be transmitted this way. As metal became available, gear makers were able to fashion smaller and lighter mechanisms for time keeping and measuring purposes. The shape of gear tooth surfaces changed along the way to triangular, then epicycloidal, and eventually the involute curve commonly used today. Wooden teeth are





**Gear Manufacturing Machines, Fig. 1** South pointing chariot

still in used today in the Duquesne Incline in Pittsburgh, Pennsylvania. Hard maple teeth with an involute profile are fitted into a rim and can be replaced as needed when they show wear or cracks.

Modern gear technology has its roots in work of mathematicians and scientists starting in the fifteenth century. Frenchman Nicholas of Cusa studied tooth profiles as early as 1451. Gear theory has been subject to near-constant study and debate ever since. As new materials and methods come into use the pioneering work of the early investigators remains worthy of study. Leonard Euler's 1754 rules for conjugate action are still valid. Edward Sang's 1852 work is still the basis of today's tooth generating machines.

Metal gears for power transmission, rather than instrument or clock movements, came into widespread use with the adoption of water power for mills. As the mills grew in size, wooden teeth were not strong enough. The introduction of the steam engine and later the internal combustion engine spurred further advances in the gear art as power density and rotational speed increased. At first, the metal teeth were cast from hand-carved patterns and hand filed to improve accuracy and reduce vibration.

As early as 1672, Englishman Robert Hooke was cutting spur gear teeth with a hand-powered wheel. By 1825, machine power was being used in Switzerland. Machined teeth became practical with advances in cutting tools, first with planing-type tools, then rotating cutters, and finally threaded "hobs." Adaptations of these methods are still in use today and will be discussed in more detail below. By 1842, Joseph Saxton of England was "generating" an involute curve profile with a straight-sided cutter.

By 1874, machine-cut bevel gears, a critical component for automotive drive trains, were available for right-angle applications. This period saw rapid advances in gear-making machinery as demand grew for better and less expensive gears for a wide range of applications. Automatic indexing, differential gearing, and tooth flank grinding were developed in the later years of the nineteenth century.

## Principle Types of Machines in Current Use

Gear-cutting machines are categorized by the type of cutting tool used and the type of gear produced. The following chart ([Table 1](#)) shows the general breakdown.

### Planing Cutters

Gear-cutting machines that use planing techniques are descended from the hand-carving methods. Planing uses a reciprocating motion to remove material from the blank and to form the desired tooth shape. The most basic machines require the cutting tool to be the shape of the space being formed; more sophisticated machines rotate the workpiece in controlled manner so as to "generate" the required curvature with a straight-sided edge. Straight cutting edges are less expensive to make and to sharpen. The process is more time consuming than milling or hobbing because only one cutting edge is in use per stroke and the reciprocating motion is not making chips during half the cycle.

### Milling Cutters

Gear-cutting machines that use rotating milling cutters are among the oldest machines still in everyday use ([Fig. 2](#)). Similar technology, although with carbide inserts bolted to a cutter body, is being used in some of the latest "gashing" machines. Rotary cutters were much faster than planers since they allowed more cutting edges to contact the blank and eliminated the non-productive return stroke. As the cutter profile must mimic the tooth space, the level of precision achievable is directly related to the accuracy of the cutter.

Gear Manufacturing Machines, Table 1 Gear type vs. manufacturing process

Type of machine	Planing	Milling	Hobbing	Shaping	Form grinding	Generating grind
Spur gears	Yes	Yes	Yes	Yes	Yes	Yes
Helical gears	Yes	Yes	Yes	Yes	Yes	Yes
Double helical gears	Yes	Yes	Yes	Yes	Yes	Yes
Internal gears	Yes	Yes	No	Yes	Yes	Yes
Racks	Yes	Yes	Yes	Yes	Yes	Yes
Straight bevel gears	Yes	Yes	No	No	Yes	Yes
Spiral bevel gears	Yes	Yes	No	No	Yes	Yes
Hypoid gears	Yes	Yes	No	No	Yes	Yes
Worms	No	Yes	No	No	Yes	No
Worm gears	No	Yes	Yes	No	Yes	Yes

## Hobbers

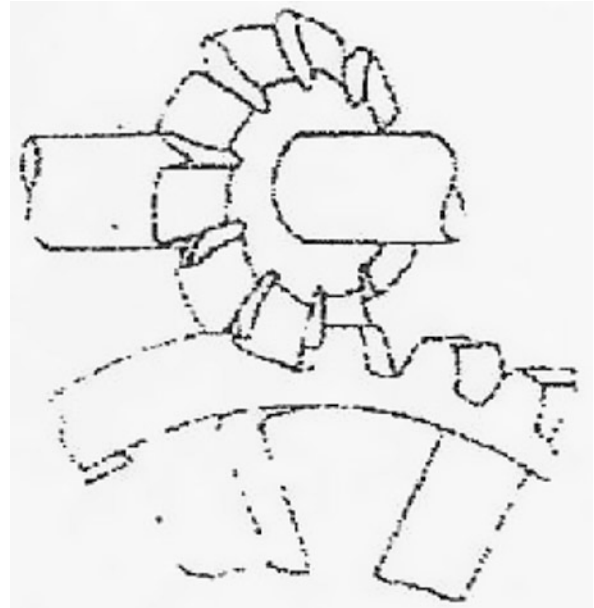
Gear-cutting machines that use hobs to cut gear teeth are by far the most common in use today (Fig. 3). The spiral nature of the cutting tool allows for many more cutting surfaces than the milling or gashing cutter. On later machines, sophisticated controls allow the operator to reposition the tool at regular intervals to use “fresh” cutting edges during the process. The machines have very long service lives – some 1940s vintage machines are still in service – so some shops require higher levels of operator skill than others.

## Shapers

Shaping involves a reciprocating pinion-like cutter that is rotated and in-fed against the rotating blank to generate the tooth profile (Fig. 4). Spur, helical, and double helical gears can be produced by this method with either internal or external teeth. Cutting tool availability is not as great as with hobbing, and machine capabilities are far more limited, especially with regard to helix angle. Special “guides” are needed for each helix angle to be cut along with matching-cutting tools. Special “herringbone” machines have been developed to cut both sides of a double helical gear at the same time with little or no “gap” between them. Larger gears (6 DP and lower) can also be produced on “rack” shaping machines that use a straight “rack” of cutter teeth to generate the tooth profile.

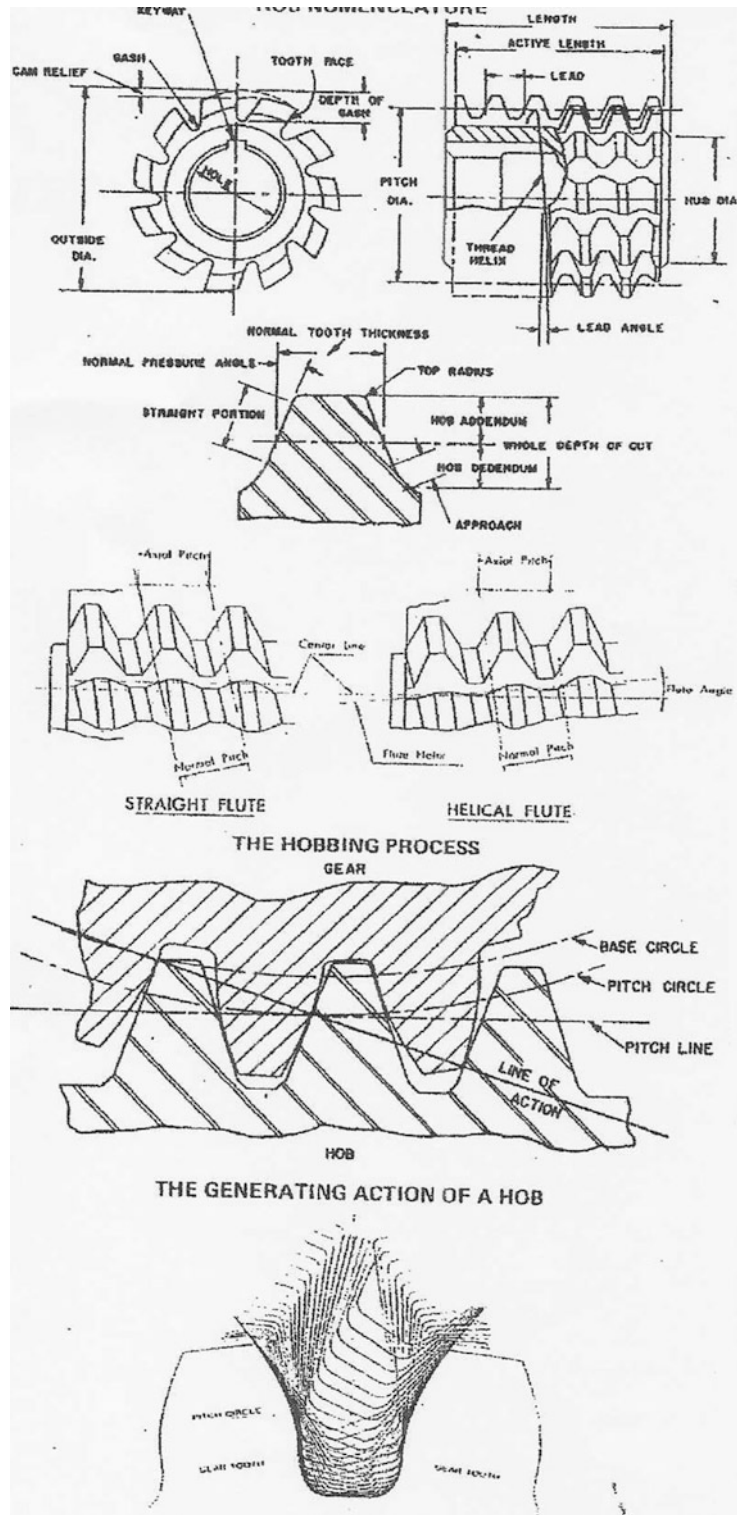
## Grinders

Gear teeth can be cut in a blank using a grinding wheel rather than a cutting tool (Fig. 5). On all but very fine pitch teeth (larger than 12 DP) this is a very



Gear Manufacturing Machines, Fig. 2 Milling cutters

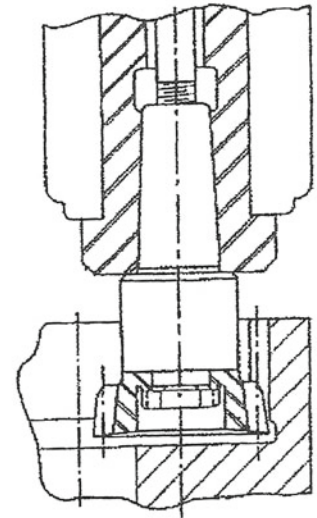
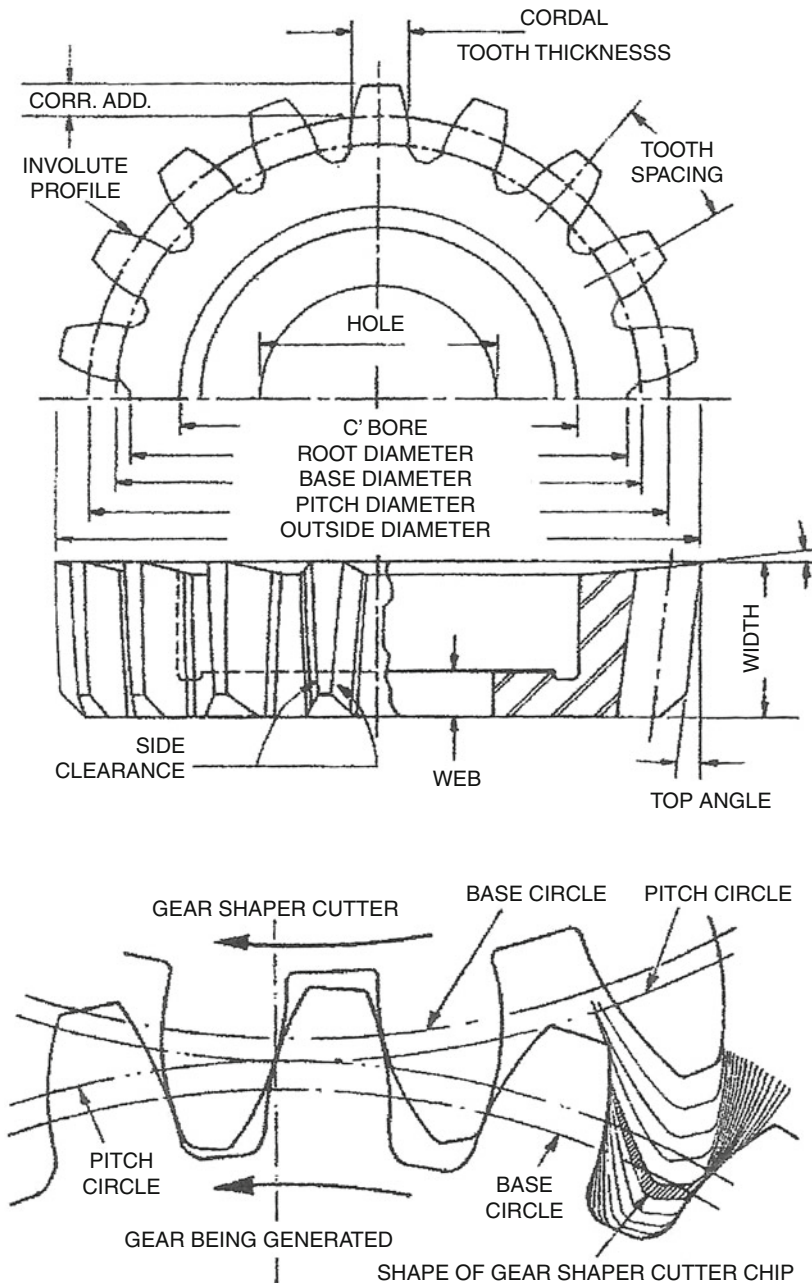
time-consuming process. As a result, it is typically avoided except for prototypes where conventional cutting tools are not available. A variety of operating principles have been applied to gear grinders and all have proven capable of producing high-quality gears. In recent years, single-wheel, form-grinding machines – frequently with generating capability – have been popular for their combination of low tool cost and high productivity. Threaded wheel type machines remain popular



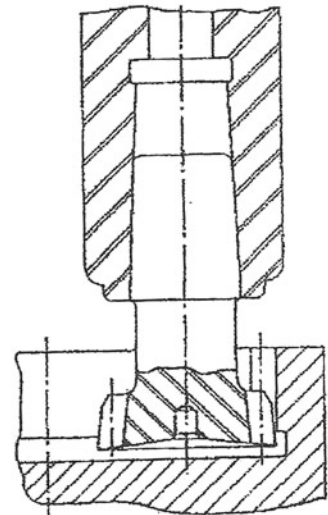
Gear Manufacturing Machines, Fig. 3 Hobbers

### Shaper Cutter Nomenclature

(Courtesy of Ash Gear and Supply Corporation)



Cutter & Holder



Shank Cutter

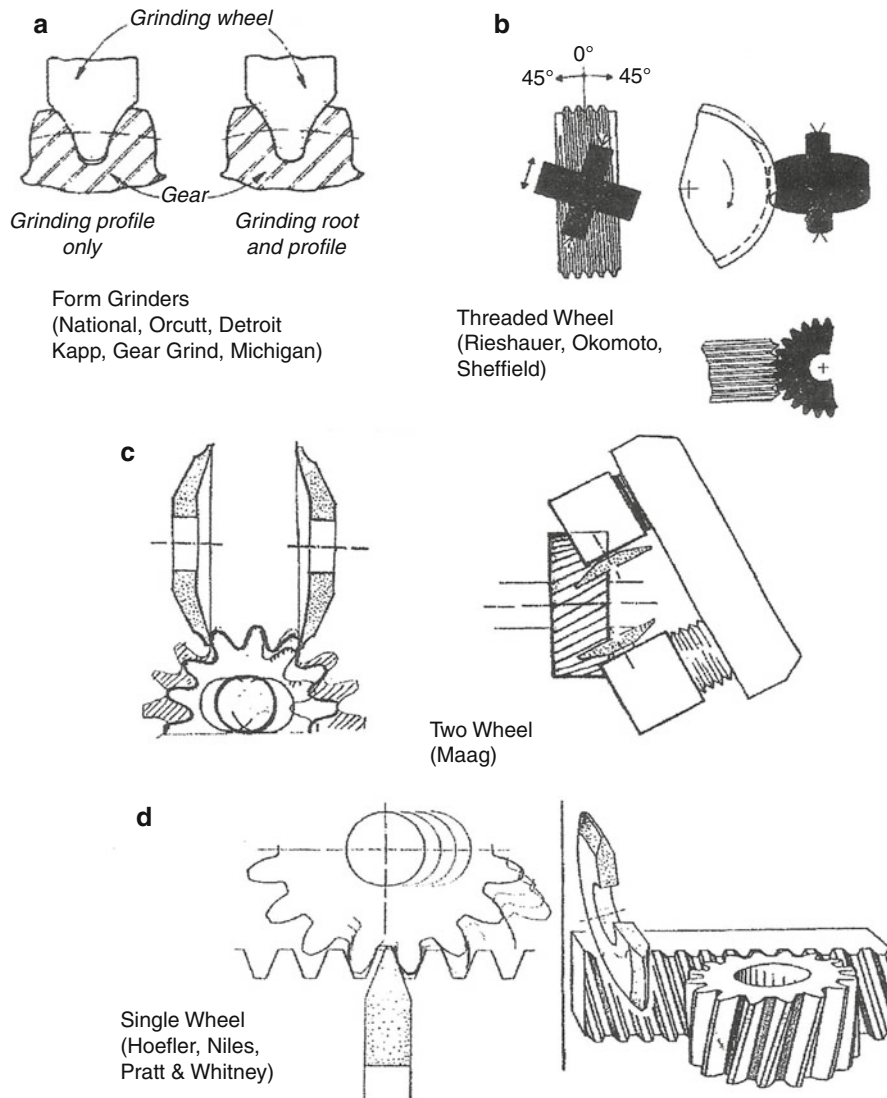
Gear Manufacturing Machines, Fig. 4 Shapers

for smaller parts, especially for high-volume applications. Technology changes rapidly in the grinding sector but the operating principles remain fairly constant. High accuracy, excellent surface finish, and special features like crowning, tooth taper, and profile modification (tip

relief) are possible with most grinding methods. While there are many different brands of grinding machines (see Fig. 5), each with a slightly different operating principle, they can be divided into two basic types – form grinders and generating grinders.



(Courtesy of American Pfauter's *Gear Process Dynamics*)



**Gear Manufacturing Machines, Fig. 5** Gear grinders types. **(a)** Form grinders. **(b–d)** Generating grinders

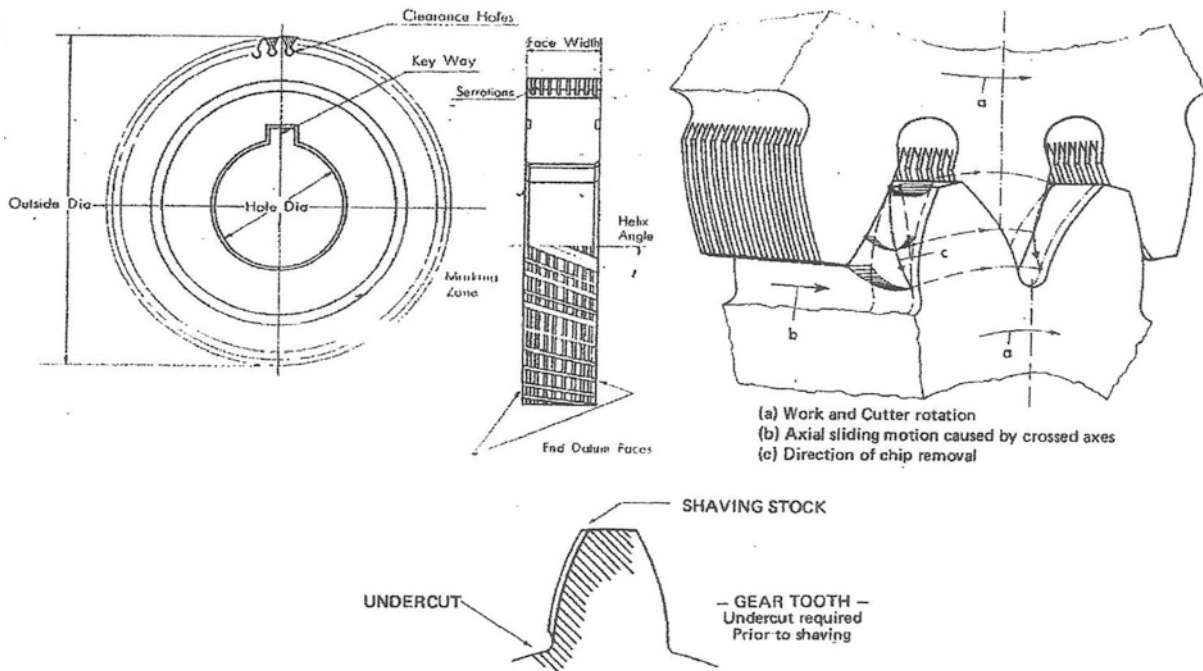
## Form Grinders

Form grinders employ a thin wheel that has been “dressed” with the profile of the desired tooth space (Fig. 5a). This method is very versatile and can be used to grind internal and external teeth of almost any type, including non-involute forms. Form grinding is also used to produce racks and gear segments that cannot be done by generating methods. Wheels as small as 1.5 in. in diameter can be fitted on some machines allowing one piece “cluster” gears and small internal gears to be processed. Newer form-grinding machines are able to use longer-lasting abrasive-coated steel wheels.

## Generating Grinders

One type of generating grinder employs a large-diameter (8–14 in. is most common) threaded wheel that acts much like an abrasive hob (Fig. b–d). A relatively simple dressing mechanism is used to maintain a very accurate straight sided rack with the proper pressure angle and depth in the wheel. Other generating grinders use either one or two narrow grinding wheels that simulate a rack.

Improper feeds, speeds, wheel materials, and coolants can cause grinding cracks and re-tempering (also called soft spots or grinding burns). Nital etching is used as an in-process check on grinding. It is also difficult to



Gear Manufacturing Machines, Fig. 6 Shaving

maintain tight controls on tooth thickness and alignments with other part features because heat treat distortion may not occur uniformly and the amount of stock removed from each side of the space may not be equal.

## Shaving

Shaving is a secondary tooth finishing process that uses a serrated cutting surface intersecting at a slight angle to the previous formed surface to improve the smoothness and the accuracy of the tooth surface (Fig. 6). Shaving is usually limited to parts of less than 400 HB.

## Broaching

The broaching process is used to produce internal spur gears and splines. A broach having the same shape as the required tooth spaces is pushed or pulled through a pilot hole with each row of teeth removing a little more metal. Some parts may require more than one “pass” with a series of broaches to reach the final size. The process is very fast and accurate but requires expensive tools and careful blank preparation. Broaching is usually limited to parts with hardness less than 45 HRC.

## Bevel Gears

Bevel gear cutting methods fall into two general categories—the reciprocating blade method for straight bevels and

rotating cutter methods for straight, spiral, and hypoid bevels. Bevel gears are probably the most difficult to manufacture. The theoretical “summary” (a computer-calculated set of machine adjustments) is just the starting point for developing the proper contact pattern between the gear and pinion. During a production run this contact pattern must be constantly monitored and adjustments made for cutter wear and diameter changes due to re-sharpening.

## Worm Gears

Worm gear sets consist of a threaded worm and a mating worm gear. The worms can be produced by rolling, milling, or grinding. Worm gears, on the other hand, can be almost as tricky to make as bevel gears. Successful development of the contact pattern and tooth thickness depends on proper initial tool design and accurate adjustment of the hobbing machine to account for cutter re-sharpening. Fly tools, commonly used on low-volume jobs due to lower tool cost, are particularly prone to this problem due to the short life of the single cutting point per lead. Worm gear hobs are custom designed to replicate the mating worm and have very distinct limits on the number of times they can be re-sharpened without compromising the tooth geometry.



## Electro Discharge Milling and Water Jet

Gear teeth can be cut using non-traditional methods such as electro discharge milling (EDM) and water jet cutting. Accuracy depends upon the sophistication of the machine's controls but AGMA Q-9 tolerances have been achieved in some cases.

## Recent Advances in Gear Cutting Machines

Computer control of machine axes has become commonplace in gear cutting machine, although the long service life of the typical machine makes it seem relatively rare. (Milling machines from the 1920s are still in use, as are hobbing machines from the 1940s.) The latest hobbbers, shapers, and gear grinders have higher power and higher speed motors on the cutter drives along with the ability to inspect the finished gear while it is still on the machine. In the 1980s, titanium nitride coatings became available on gear-cutting tools, greatly increasing tool life while simultaneously increasing productivity. Cutting tool technology, particularly with regard to carbide inserts, continues to allow higher levels of productivity if machine rigidity and sufficient power are available.

## Cross-References

- Gear Cutting Tools
- Gear Manufacturing
- Gear Surface Treatment

## References

- D.W. Dudley, *The Evolution of the Gear Art*. American Gear Manufacturers Association Library of Congress Catalog Card Number: 72-78509
- D.W. Dudley, *Practical Gear Design*. McGraw-Hill Library of Congress Catalog Card Number: 53-11476
- D.W. Dudley, *Gear Handbook*. McGraw-Hill Library of Congress Catalog Card Number: 61-7304
- C.D. Schultz, *An Introduction to Gear Design*. Privately published 1987-1997. Available for free download at [www.beytagear.com](http://www.beytagear.com)

---

## Gear Materials

PHILIP TERRY

Chief Metallurgist – P Terry and Associates, Lufkin, TX, USA

## Definition

This section presents a review of the wide range of materials used in industrial gears, including steels, irons and non-ferrous alloys.

## Scientific Fundamentals

Gear materials are subjected to multiple load cycles during operation, and gear teeth experience high surface (Hertzian) contact stresses, high bending stresses, typically in the root area, and potential sliding wear forces as the teeth move in and out of mesh. The contact stresses and bending stresses produce long-term fatigue damage. Many design standards exist to estimate the life of gear teeth from these two fatigue mechanisms for various applications and when manufactured from various materials.

More information is available in standards such as American Gear Manufacturers Association (AGMA) 2001 (ANSI/AGMA 2001 D04 Fundamental Rating Factors and Calculation Methods for Involute Spur and Helical Gear Teeth). In the following, reference will be made to the allowable contact stress  $S_{ac}$  and allowable bending stress  $S_{at}$  from AGMA 2001 when comparing different gear materials.

## Key Applications

### Steel Gear Materials

Steel gearing is manufactured from a broad range of materials, from basic carbon steels of the SAE (Society of Automotive Engineers (SAE) SAE HS-1086 Metals and Alloys in the Unified Numbering System) 1045 type to alloy steels containing significant additions of nickel, chromium, and molybdenum, such as the SAE 9310 carburizing alloy.

Steel gears for larger parts are produced by common manufacturing methods, including casting, forging, and ring rolling. Small gears, typically in the range of 100-mm diameter or less, are also produced by stamping and fine blanking, and high volume small gears are manufactured by powder metallurgy (ANSI/AGMA AGMA 6008 Specifications for Powder Metallurgy Gears) processes.

Cast gears, which can be up to 18 m in diameter for applications such as girth gears for kilns and SAG mills, are used in the as-cast condition or after heat treatment, including normalizing, normalizing and tempering, quenching and tempering, and carburizing. The range of alloys used in cast steel gearing includes SAE 1045, 4140, 8630, 8642, and 4340 type compositions. Additional information on cast steels may be found in publications from the Steel Founders' Society of America (SFGSA) (Steel Castings Handbook) and ASM-International (ASM Handbook, Volume 15, *Casting*). As with all steels, the selection of the appropriate alloy is based on required mechanical properties in the finished part, cross section (dimensions controlling heat treatment response), and restrictions imposed by other processing. Such process-related

restrictions include analysis restrictions imposed by processes such as carburizing or requirements for control of hardenability for parts that are welded.

Forged steels, those that are deformed at high temperatures to consolidate the metallurgical structure, produce fine grain size, and optimum properties are used to manufacture gears from several initial forged shapes. Near net shaped forgings are used in the manufacture of large integral gears for applications such as power generation. Large, flat “pancake” shaped forgings are used to produce gears that are fitted or keyed to separate shafts. Forged rounds are machined to produce similar geometries for smaller size gears, and rolled alloy steel rings are used in large fabricated gears for applications such as steel mills and sugar mills. High-quality rolled steel rounds are available, but the diameter at which overall soundness – freedom from central imperfections – can be achieved is limited, and the size of gearing that is permitted to be machined from rolled steel rounds is often restricted.

The range of steels used for forged gear manufacture is typically the higher alloy SAE 4140, 4340, 4350 grades, which receive heat treatments including normalize, normalize and temper, and quench and temper as applied to cast steel products. Again, the selection of the appropriate alloy is largely based on required mechanical properties in the finished part and cross section.

Requirements for material quality of cast steel and forged steel parts are included to some degree in a number of individual gear application standards, such as those for vehicle gearing, gears for marine applications, and some general requirements in the fundamental rating standard AGMA 2001. Efforts have been made both within the AGMA and the International Standards Organization (ISO) to consolidate such requirements, and the reader is referred to AGMA 923 (AGMA 923 Metallurgical Specifications for Steel Gearing) and ISO 6336-5 (ISO 6336-5 Calculation of Load Capacity of Spur and Helical Gears – Part 5: Strength and Quality of Materials) for additional comprehensive information on the metallurgical requirements that need to be applied to the various material process routes and product forms used for gears. Some specialized applications, such as aerospace gears, demand more enhanced quality requirements, beyond AGMA 923 and ISO 6336, and many individual companies have developed extensive experience in specific areas and specify additional requirements that they have found enable them to operate gears more reliably.

While many cast and forged steels are subjected to bulk heat treatments such as normalizing or quenching and tempering in order to develop the necessary properties to resist the loads imposed on the gear teeth, surface

enhancement techniques to improve resistance, particularly associated with gear tooth surface stresses, are also employed. These include flame and induction hardening, nitriding, carburizing, and nitrocarburizing.

Flame and induction hardening are performed on the same range of alloy steels used for cast or forged steel parts. The core material is heat treated to the required strength level and then the tooth surface is hardened and tempered at low temperature to a surface hardness in the range of 50–55 Rc.

The nitriding process is also carried out on the same range of alloys, which again are bulk heat treated, typically by quenching and tempering prior to the nitriding operation. The nitriding process is performed at relatively low temperatures – around 900°F – and causes little if any distortion of the part, making it applicable to complex-shaped parts, thin section ring gears, and similar geometry components that would change shape to an unacceptable degree if subjected to other processes. Because the nitriding process is performed at low temperatures, however, diffusion rates for the process are very slow and case depths are usually limited to a maximum of around 0.030 in. Surface hardness levels on the nitrided case are around 83.5 HR15N (approximately 45 Rc) on conventional alloy steels; special nitriding alloys with deliberate additions of aluminum, chromium, and other alloys are also available and these can develop surface hardness levels of up to 90 HR15N (approximately 60 Rc). Nitrided gearing is particularly suited for applications where surface damage mechanisms such as scuffing are an issue.

Carburizing is performed at high temperatures where the steel material is in the austenitic high temperature form of iron. Austenite has a high solubility for carbon and diffusion rates are high – deep cases can be produced in relatively short time cycles and the final product is effectively a composite structure with a low carbon alloy steel core and a high carbon material on the tooth surfaces. After carburizing, the component is quenched, which produces a low hardness ductile core and a high hardness tooth surface. The metallurgical changes occurring during the quenching process are also modified by the presence of the high carbon level in the case. The high carbon case transforms later than the low carbon core, and this results in the development of significant compressive residual stresses in the surface. After final low temperature tempering around 350°F to restore ductility to the as-quenched material, the part possesses both high hardness surfaces and also surfaces with high levels of residual stress, which in combination give carburized gear teeth the highest load-carrying capacity of all the

material and processing options available. In order to achieve this optimum combination of properties, special steel is used for carburized parts. Typically carbon contents in the initial material do not exceed 0.2% because above this level the final core properties are too brittle. In addition to control of the carbon content, in order to ensure that the core material is sufficiently ductile, even following a low-temperature temper carburizing, materials often contain significant alloy additions of nickel. Alloys used for carburized gearing range from SAE 8620 with 0.5% nickel to SAE 9310 with 3.5% nickel. Other alloying elements, including chromium and molybdenum, are also added to carburizing steels in order to enable them to develop the required core mechanical properties. Guidance on alloy selection for various tooth sizes and various cross sections of material is offered in AGMA 2004 (AGMA 2004 Gear Materials. Heat Treating and Processing Manual).

As a function of the thermal cycles to which parts are subjected, carburizing results in significant shape changes and the parts require post-carburizing machining and grinding to correct the shape changes and produce accurate tooth profiles and surface finish. The sizes of available grinding machines and the limitations on the amount of post-carburizing grind stock that can be put on parts restricts the maximum size of carburized gears. Depending on tooth size, carburized gears are limited to around 60-in. diameter for finer pitch gearing of 3.5 NDP and less, while coarser pitch teeth gears up to 100-in. are produced.

Nitrocarburizing is used on lower cost, lean, low alloy steels to produce a hard but very thin surface. This process is used for fine pitch teeth on mass-produced parts.

There are a great many variables to take into account when comparing the allowable stress rating numbers for through-hardened gears – hardness, quality levels, and so on – and even more for case-hardened gears, but a general ranking order would be as follows:

Allowable Contact Stress from Lowest to Highest:

- Through-hardened steel
- Nitrided steel
- Flame- or induction-hardened steel
- Carburized steel

Allowable Bending Stress from Lowest to Highest:

- For flame- or induction-hardened steels, there is pronounced influence of the shape of the hardening profile on the tooth. Hardening patterns that do not include the tooth root actually have the lowest bending strength.

- Flame- or induction-hardened gears with hardening profiles that include the root, through-hardened gears, and nitrided gears have broadly similar bending strength.
- Carburized gear teeth, as with contact stress, have the highest known load-carrying capability against bending.

Overall, carburized and hardened gear teeth have the highest load-carrying capacity. In order to fully utilize these properties, a high level of control of the raw material and processing is required, as examination of the various AGMA and ISO documents reveals.

### Cast Irons

Cast irons used in gearing fall into three main groups: gray cast iron, nodular (or ductile) iron, and austempered ductile iron.

Gray cast iron, typically to ASTM A 48 (ASTM A 48 Specification for Gray Iron Castings) or similar national standards, contains approximately 2% carbon, and during solidification the carbon forms graphite flakes within the material. When broken, the material breaks along the weak graphite and the graphite gives a the fracture surface a gray appearance, thus “gray cast iron”.

The size and distribution of the graphite flakes within the iron matrix significantly affect the mechanical properties of the material; shorter, more randomly distributed graphite result in increased strength. Graphite morphology is therefore often specified in addition to basic strength requirements.

Of all the types of cast irons available, gray iron is the easiest and lowest-cost material to produce. It has excellent mechanical damping characteristics and good thermal conductivity, however, the fatigue strength properties of gray cast iron are low both for gear tooth pitting resistance and gear tooth bending strength, with absolute maximum allowable value of only 85,000 psi  $S_{ac}$  and 13,000 psi  $S_{at}$  in AGMA 2001.

Ductile iron is a modified form of cast iron, generally of similar composition to gray cast iron but with the addition of magnesium or cerium bearing alloy, which modifies the surface tension between the graphite and the molten iron, causing the graphite to form nodules or spheres, as opposed to flakes. The formation of the graphite as nodules and greater separation of the graphite within the iron matrix produce a significant increase in properties, most dramatically in ductility, from which the material derives its name. ASTM has five standards covering ductile iron castings. ASTM A 536 (ASTM A 536 Specification for Ductile Iron Castings) is the most frequently used, and within this

standard the 100-70-03 grade is most commonly employed for gearing. The three-number designation system indicates tensile strength, yield strength (both in units of ksi), and percentage elongation.

Ductile iron is more expensive to produce than gray cast iron, requiring both the addition of magnesium or cerium alloy and closer control of overall chemical analysis of the melt and tighter controls of temperature during teeming. Ductile irons are also more difficult to machine, have lower damping capacity and lower thermal conductivity than gray cast irons, but the superior mechanical properties and, in particular, the dramatic improvement in toughness, lead to significantly greater use of ductile iron gearing compared with gray iron.

Ductile irons are used for gearing in the as-cast condition and also after heat treatments such as quenching and tempering, normalizing and tempering, and the specialized austempering process, described later. Ductile iron gearing is also often surface-hardened by flame- or induction-hardening techniques. The ASTM A 536 Grade 100-70-03 material, for example, is usually produced by normalizing and tempering, while lower strength grades can be produced as-cast.

The ASTM A 536 100-70-03 grade of ductile iron has maximum allowable stress levels in AGMA 2001 of 112,000-psi  $S_{ac}$  and 40,000-psi  $S_{at}$ .

Austempered ductile iron (ADI) is produced by heat treating cast ductile iron. After austenitizing, the material is controlled cooled to produce a structure of ferrite, or ferrite plus carbide, in a matrix of high carbon austenite. This microstructure results in a material superior to cast ductile iron, particularly in strength and toughness. The various strength level grades, quality, and related requirements for ADI are contained in ASTM A 897 (ASTM A897 Standard Specification for Austempered Ductile Iron Castings). There are no formal rating numbers for ADI, but the information sheet AGMA 939 (AGMA 939 Austempered Ductile Iron for Gears) contains some general guidance. While ADI material is generally considered to have overall superior properties compared with gray cast iron and cast ductile iron, there are some issues regarding its use in gearing. Machining of ADI material requires special techniques, tool selection, and feeds and speeds. Other cast iron and steel gear materials show an increasing resistance to both Hertzian contact and bending stresses with increasing hardness. In the case of austempered ductile iron, this remains true for resistance to contact stress but there is an inverse relationship between fatigue life and hardness for bending strength rating, and this unusual behavior needs to be examined when considering the use of ADI in gears.

In order to calculate gear tooth life using the allowable stresses in the various national and international standards, specific minimum material quality levels must be met. These quality requirements are detailed in documents such as AGMA 923, which covers steel gearing, and ISO 6336-5, which, in addition to steel gearing requirements similar to AGMA 923 also covers metallurgical quality requirements for conventional gray and ductile irons.

## Non-ferrous Gear Materials

The most significant non-ferrous metallic alloys used in gearing in terms of volume are copper-based alloys and, in particular, the alloy range known as bronzes. Bronzes used in gearing are mainly the tin-bronzes typified by the Unified Numbering System (UNS) alloy C90700, often called phosphor gear bronze. The tin-bronze alloys are tough and have good corrosion resistance. They also possess good sliding wear characteristics, which is why they are often used in spiral bevel gears, worm wheels, and other severe wear applications.

Manganese bronzes, either cast, such as UNS C86200, or wrought, such as UNS C65700, are used. These alloys have good wear resistance but are not as corrosion resistant as the phosphor bronze or aluminum bronzes. Aluminum bronzes can be heat treated to higher strengths than the manganese bronze alloys and can be used to produce lighter-weight gearing. They have good wear resistance, a low coefficient of friction against steel, and bearing characteristics between those of the manganese and tin bronzes. Typical aluminum bronze gear alloys are UNS C95200 casting alloy or UNS C62300, which is used in the wrought condition.

Silicon bronzes, such as cast UNS C87600 or wrought UNS C65500, are used in lightly loaded gearing for electrical applications because of their low cost and nonmagnetic properties.

Common yellow brass and similar copper-zinc alloys are used for small complex gearing because of ease of machining, even though the wear resistance of the brasses is not as good as the bronze alloys.

Users need to be aware, however, of the potential for corrosion damage to copper-based alloys due to the effects of extreme pressure additives in some lubricant formulations, particularly those utilizing sulfur compounds.

Standardized guidance on allowable contact and bending stresses for bronzes are limited. AGMA 2001 contains some information, but generally  $S_{ac}$  and  $S_{at}$  values for bronzes are low and, hence, the use of these alloys is usually restricted to those areas where the sliding wear characteristics of the alloys are the overriding design requirement.

Aluminum alloys are also occasionally used for gearing, but applications are so specialized that no general guidelines on these alloys are published. Selection of material and tooth stress calculations are each usually resolved on a case-by-case basis between the gear designer/manufacturer and the end-user.

### Non-metallic Gear Materials

The topic here is non-ferrous metallic materials, but it should be noted that various non-metallic materials are also used in gearing. Wooden gear teeth are still in operation on the Duquesne Incline in Pittsburgh (Duquesne Incline, see [www.duquesneincline.org](http://www.duquesneincline.org)), and the use of all metals in gears is dwarfed by plastics, in terms of numbers of gears used on a daily basis. Plastic gears see universal use in office machines and motor vehicles. Details of requirements for plastic gearing materials are covered in AGMA 920 (AGMA 920 Material for Plastic Gears).

### Cross-References

- [Gear Bending Fatigue Failure and Bending Life Analysis](#)
- [Gear Contact Temperature and Scuffing Risk Analysis](#)
- [Gear Surface Pitting Failure and Pitting Life Analysis](#)

---

## Gear Molding

- [Gear Manufacturing](#)

---

## Gear Noise

DONALD R. HOUSER

Gear and Power Transmission Research Laboratory,  
Department of Mechanical and Aerospace Engineering,  
The Ohio State University, Columbus, OH, USA

### Synonyms

[Gear dynamics and noise](#); [Gear rattle](#); [Gear vibration and noise](#); [Gear whine](#)

### Definition

Whenever gears rotate at substantial speed, they will generate noises that are often objectionable and are sometimes at levels that are dangerous to the listener's

hearing. The most common complaints about gear noise come from transportation and consumer products such as motor vehicles, aircraft, ships, appliances, printers, and hand tools. Noise is generated from motion error at the mesh and also due to deflections occurring due to stiffness changes within the mesh. Although there are several types of gear noise, this discussion focuses on gear whine noise and provides discussions of the tribological influences on gear noise.

### Scientific Fundamentals

Three types of noises are typically produced by gears:

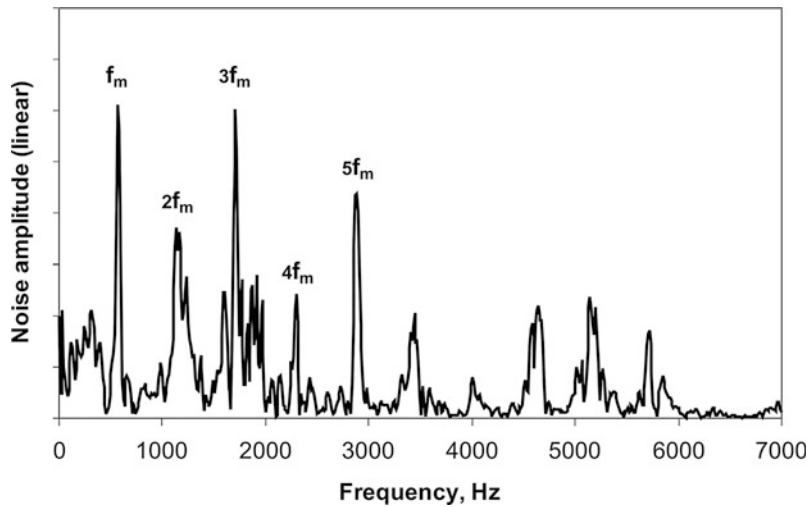
***Gear Whine:*** This sound is tonal in character and typically occurs at the gear mesh frequency and its harmonics. Most of the discussion here will be devoted to this type of sound. The major periodicity of gear whine is at the gear mesh frequency that is calculated by multiplying the number of teeth on the part by its rotational speed. [Figure 1](#) shows the frequency spectrum of the noise from an FZG gear tester that is commonly used for gear durability and gear lubricant testing. One sees not only sounds at the mesh frequency,  $f_m$ , but also at harmonics of mesh frequency ( $2f_m$ ,  $3f_m$ ,  $4f_m$ , etc.).

***Gear Rattle:*** This sound is most common in lightly loaded gears and results from a train of repeated impacts. Gear rattle does not have any individual frequencies but is broadband in nature. Gear rattle is normally caused by an external excitation and will only briefly be covered in this chapter.

***Gimmick Sounds:*** Interestingly, gearboxes can generate sounds that are neither of the above and are unique in their signature. Examples include clunks, plastic gear screeching, clicks due to nicks, and so on. Since each gimmick sound is unique and is quite often unrelated to the gears themselves, reduction of the sound requires a unique solution so these sounds will not be covered here.

Many detailed summary articles (Houser [2007](#); Singh [2009](#); Welbourn [1979](#); Singh and Houser [2012](#)) have been written on gear noise, so this discussion will present a broad overview of the topic and will focus on spur and helical gears, but the principles presented are appropriate to all gear types. The basic differences of gear types can be reduced to considering when the gear type has conjugate motion. The involute tooth form that is used for spur and helical gears provides perfect gear motion transfer (conjugacy) across a range of center distances, while in most other gear types such as bevel, hypoid, or worm gears, conjugacy occurs at only one center distance.





**Gear Noise, Fig. 1** Frequency spectrum of gear noise emanating from an FZG gear testing machine

Hence, misalignments and other effects that tend to change center distance will create greater motion errors in these gear types.

### Gear Noise Generation

Gear noises originate at the gear teeth, where force and motion are transferred from one gear to its mating gear. The mesh force is time varying with the periodicity occurring at the gear mesh frequency. Here, the examples will show results in terms of the input or output amplitude at the first mesh harmonic,  $f_m$ , however, it is typical to have sounds at all integer harmonics of mesh frequency as shown in Fig. 1. Other periodicities might occur due to load and driver pulsations with these frequencies being at multiples of their respective shaft rotational frequencies. A depiction of the process of noise generation in the classic source-path-receiver format is shown in Fig. 2. Although the block diagram applies to all types of gear noise, the following discussion is mainly relevant to gear whine noise.

### Mesh Excitation

The main sources of gear whine excitations are:

**Transmission error and mesh stiffness variation:** These two excitations go hand in hand. Transmission error (TE), which is described in the chapter of this handbook entitled Gear Dynamic Loading, is the rotational motion error of the gear pair (often referred to as non-conjugacy). For lightly loaded gears, transmission error is due to deviations of the tooth surface from a true involute, and for loaded gears one must add the deflections of the tooth pairs due to stiffness changes in the mesh. In the simplest

model for gear dynamics, shown in Fig. 3, transmission error acts as a displacement input depicted at the cam and stiffness variation that occurs at the mesh spring.

$K_b$  and  $k_b$  are bearing/shaft stiffnesses.

$C_b$  and  $c_b$  are bearing/shaft viscous dampers.

$k_m$  and  $c_m$  are the respective stiffness and damping of the gear teeth.

TE is transmission error.

$R_b$  and  $r_b$  are the base radii of the respective gears.

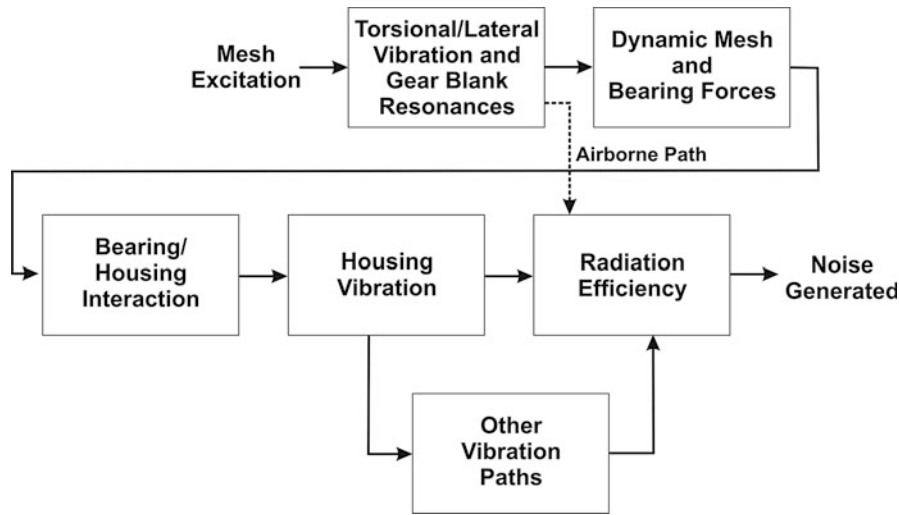
$\theta_1$  and  $\theta_2$  are the rotary motions of the respective gears.

**Friction:** Both Coulomb friction and viscous damping exist in the gear mesh. In the dynamic model of Fig. 3, viscous damping along the line of action (line tangent to the respective base circles) is depicted by  $c_m$ , while sliding friction, which acts at right angles to the line of action, would require adding an additional degree of freedom for each gear. Obviously, friction excitations are heavily related to gear tribology since the lubricant viscosity and surface finish directly impact the coefficient of friction as well as the viscous damping.

**Tooth force axial shuttling:** This is an excitation that mainly exists in helical gears where the centroid of the tooth force shifts back and forth axially along the face width, thus causing a rocking excitation to the gear and a time varying force excitation at the bearings. This force was initially studied by Börner and Houser (1996).

**Lubricant and air entrapment:** As gears rotate, there is an action to attempt to squeeze air and lubricant out of the clearance cavity at the root of the tooth. In high-speed gearing in particular, there may be a whistling sound due to this action. However, in most common gear



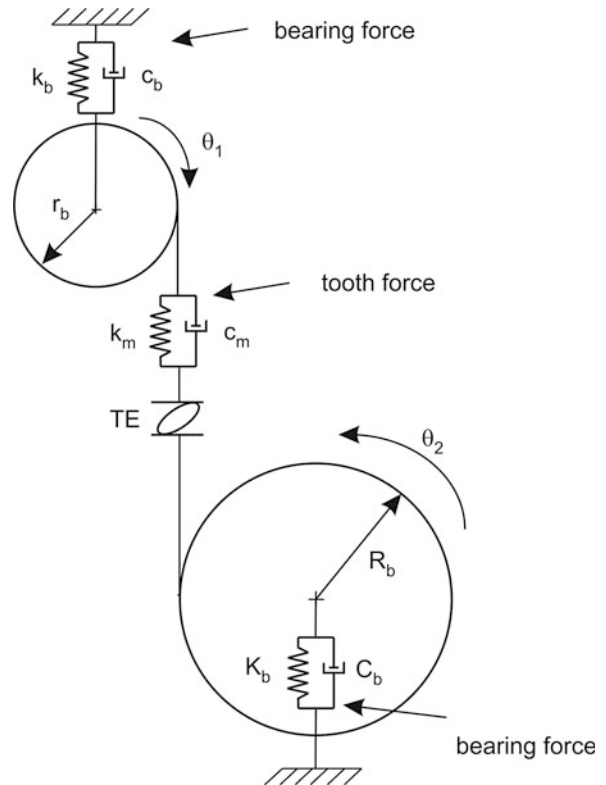


Gear Noise, Fig. 2 Source-path-receiver diagram of gear noise generation

applications, the noises generated by entrapment are usually considered to be negligible relative to the three sources that have already been discussed.

For the comparison of mesh excitations, two gear pairs, one a spur gear pair and the other a helical gear pair, have been modeled using the Load Distribution Program (LDP) (Houser and Harianto 2009). The gear pairs are two of seven pairs that were noise tested at NASA and whose results were subsequently compared with transmission error modeling (Houser et al. 1994). The seven gear pairs cover a broad range of contact ratios and the analysis results for the lowest contact ratio (CR) spur gear pair (CR = 1.43) and the highest contact ratio helical gear pair (CR = 4.37) are presented here. Each of the gear pairs was analyzed over a broad range of torques, both for a perfect involute tooth shape and for a shape in which tooth modifications were applied such that the tooth pair had minimum mesh frequency transmission error at a torque of 400 Nm. The topographical modification applied to the helical pinion in order to reduce transmission error is shown in Fig. 4 (a similar modification was applied to the example spur pinion). The axis of the figure labeled roll angle depicts the modification in the tooth form from near the root of the tooth (low roll angle) to the tip of the tooth (high roll angle).

Figure 5 shows the transmission error mesh force (computed by multiplying the transmission error by the average mesh stiffness) for the four example cases. It is evident that for perfect involutes, the transmission error force increases linearly with torque and that the transmission error force is much less for the helical gear pair than it



Gear Noise, Fig. 3 A simple model of mesh dynamics

is for the spur gear pair. The second observation is that the microgeometries (combination of tip relief and lead crown) used to minimize transmission error at 400 Nm torque do indeed cause a reduction in transmission error

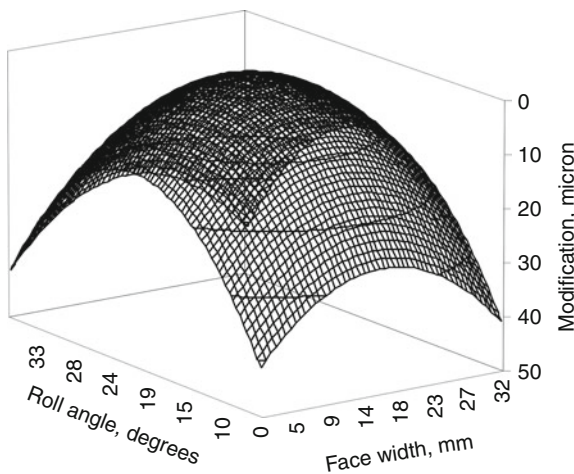
force at that torque, but for the spur gear pair low transmission errors only occur near the design torque while there is a reduction across a much broader range of torques for the helical gear pair. These are major reasons why helical gears are used in noise-sensitive applications such as automotive gearboxes and ship transmissions.

Figure 6 shows the effective bearing forces due to transmission error, shuttling, and friction, respectively, for the helical gear pair. The transmission error force graph is an expansion of this gear pair's plot shown in Fig. 5. First, it is interesting that both the transmission error force and the shuttling force have multiple torques

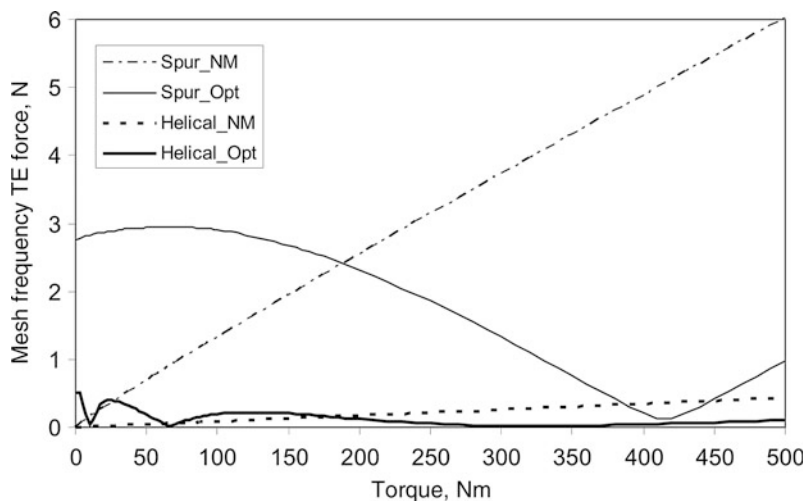
for which they go to a minimum. The number of these minima is typically one less than the integer of the contact ratio. Friction force, the factor most influenced by tribological considerations, is usually a lesser contributor, with the exception being when the other forces are near their minimum values in which case friction can be significant. Here, the only torque where friction is significant contributor to the total force is at about 370 Nm. The total force shown in the figure is computed by simply adding the effects of the individual components. In this example, friction force was computed using a coefficient of friction of 0.05, a typical value for lubricated gears. A further discussion of the concepts discussed in the previous paragraphs is presented by Houser and Harianto (2005).

### Mesh Dynamics Blocks

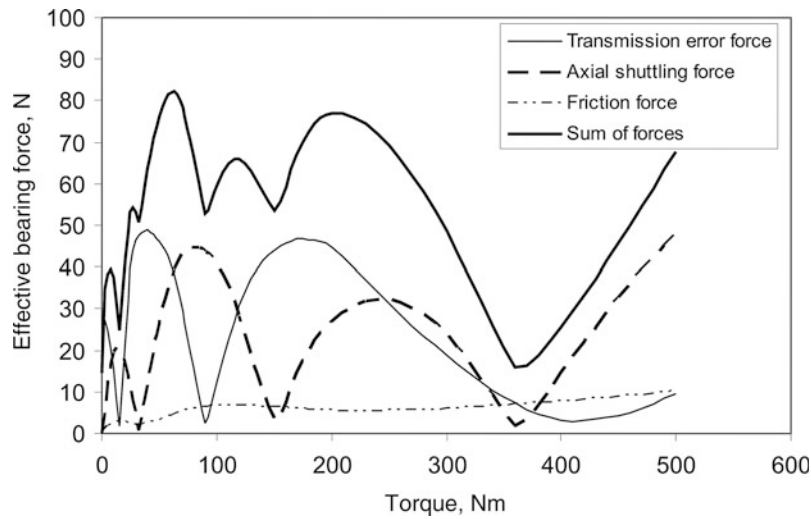
The second and third blocks in Fig. 2 may be lumped together as mesh dynamics blocks since each has an effect on the tooth forces at the mesh and hence, the forces transmitted through the bearings to the housing. Models for mesh dynamics have been treated extensively in the literature (Singh 2009; Ozguven and Houser 1988). Figure 3 is the simplest model for evaluating gear noise excitation effects since it not only has mesh dynamics incorporated via the gear masses and rotary inertias, mesh stiffness, mesh damping and transmission error, but also includes the effects of bearing and shaft stiffnesses as well as bearing damping. From a tribology standpoint, each of the viscous dampers is affected by the lubricant selection.



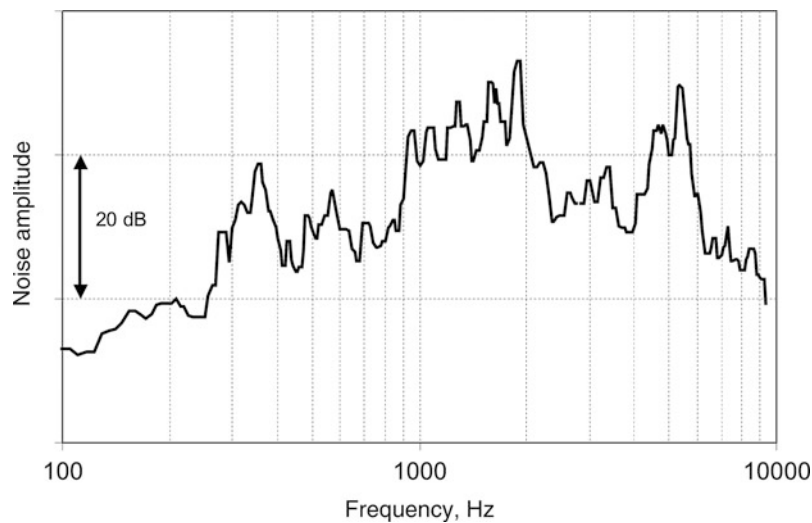
**Gear Noise, Fig. 4** Topographical modification applied to each helical pinion tooth to minimize transmission error



**Gear Noise, Fig. 5** Effect of torque on the TE force of two example spur and helical gear pairs, both as pure involutes and with microgeometry modifications



Gear Noise, Fig. 6 Comparison of the bearing forces due to transmission error, friction and shuttling for the helical gear pair



Gear Noise, Fig. 7 Influence of increasing shaft speed (mesh frequency) on gear noise

### Housing Vibration and Noise Transmission Paths

The final blocks of the block diagram can be lumped as transmission paths to the receiver and for modeling purposes, require advanced vibration and noise analyses that are beyond the scope of this chapter. A summary of some of the analyses that have been used has been presented by Singh (2009). From a tribological viewpoint, the lubricant is likely to have little effect on the noise and vibration transmission through the housing.

Figure 7 shows the mesh frequency sound generated by a pair of FZG test gears during a speed run-up. Since the load is held constant in this test, this trace is essentially the effect of all blocks in Fig. 2 while holding the excitation constant. In the low-frequency region, the noise increases at a rate of roughly 8 dB per doubling of frequency. Other experiments in different gear applications have shown that the rise in noise with frequency could range from 6 to 12 dB per doubling of frequency. Each mini-peak is a local resonant frequency whose peak level can be reduced by

adding damping. This is interesting, since here, adding energy dissipation helps reduce noise, but at the excitation, adding dissipation would have the effect of increasing the friction force and hence, increasing noise.

### Approaches to Noise Reduction

In reducing the noise of a gear pair, one may achieve improvements by working on any of the elements of Fig. 2. In more general terms, this means either reducing the source excitation, reducing transmission path effects, or modifying the receiver's environment. If one wishes to reduce the source excitation, changing the gear macro-geometry will have the greatest effect, but surface finish and lubricant may also be influential. If resonances such as those shown in Fig. 7 contribute to the noise, certainly, working on the mesh dynamics and path dynamics may provide noise reductions at the specific resonant frequencies, but will not change the overall shape of the noise trace shown in the figure.

### Gear Geometry

As hinted in Fig. 5, increasing the contact ratio (average number of teeth in contact) is the major way a gear designer can reduce gear noise at the source. Generally, increasing contact ratio is achieved by increasing helix angle, having longer teeth, reducing pressure angle, and increasing the face width of helical gears. Increasing the number of teeth by reducing module has mixed effects. On one hand, reducing module allows one to have a higher contact ratio, but, secondly, mesh frequency increases, so if one is operating in the lower frequency region, as shown in Fig. 7, noise will go up.

### Gear Microgeometry

Gear microgeometry covers any attempt to modify the shape of the tooth surface to compensate for deflections or to adjust for manufacturing and assembly errors such as misalignment. These modifications, a typical example being shown in Fig. 4, are typically referred to as tip or root relief when applied in the tooth profile direction and as lead crown or lead end relief when applied in the face width direction. Figure 5 shows that these modifications can significantly reduce transmission error and, hence, gear noise. The same microgeometry modifications used for noise reduction usually reduce scoring and micropitting, two common tribology-influenced gear failure modes.

### Gear and Assembly Accuracy

Typical accuracy measures include deviations from the desired microgeometry, spacing errors, run-out,

misalignments, and so on. Provided the basic gear design is made with the incorporation of proper noise reduction concepts, inaccuracies will typically result in noisier gears. However, if the baseline design is inherently noisy, inaccuracy will have a lesser effect.

### Tribology

The selection of lubricant, lubricant temperature, and additives probably has a secondary role in gear noise generation, but for gears designed to be quiet by having low transmission error, lubricant effects are likely to be more significant. As discussed earlier, reducing the coefficient of friction reduces friction forces transmitted to the housing. One means of reducing the coefficient of friction is to reduce the surface finish of the gear tooth surfaces. In one such instance, a combination of surface finish enhancement and microgeometry modification reduced the noise of helicopter spur gears by 7 dB (Hansen et al. 2006). The fact that the sound from automotive gears sometimes changes with the temperature of the lubricant also indicates that lubricant viscosity plays a role in gear noise generation. Finally, many gear pairs have an increasing noise trend as their load torque increases, independent of whether transmission error is increasing or decreasing. Since friction is a force that continually increases with load, as does shuttling force, this adds further speculation that tribological effects influence gear noise levels.

### Gear Rattle

Gear rattle usually occurs at light loads and is often excited by torsional pulsations coming from either the prime mover or the load device. Rattle is common in manual transmissions of vehicles, where the torque fluctuations of the engine are transmitted through the clutch to the transmission gears. Reduction of rattle noises is usually accomplished through proper clutch design (Singh 2009) but may also be influenced by the transmission lubricant. In engine balance gears, where there is virtually no load on the gears, one approach to rattle reduction is to keep the gears as conjugate as possible so the lubricant tends to use a broad surface area to "cushion" the impact of the teeth.

### Effects of Tribological Related Failures on Gear Noise

Scoring, pitting, and micropitting are gear failure modes that are heavily influenced by the lubricant and its application. When these failure modes are progressing, they are changing the microgeometry of the gear tooth and, hence, are changing the transmission error. Both scoring and micropitting tend to occur on all teeth,

so they are more prone to increasing mesh frequency noises as the transmission error increases. Pitting failures that are randomly located around the gear will increase broadband noise but will have a lesser effect on mesh frequency tones. For gears that are initially quiet, the onset of any of these failure modes is likely to have a resulting increase in gear noise.

## Key Applications

A brief summary has been presented of gear noise generation mechanisms and has discussed the impacts of tribology on gear noise. In general, tribological considerations have a secondary role, with proper macro- and microgeometry design having a major effect on gear noise. However, when these design considerations are done properly, the influence of lubrication becomes a more important factor in the generation of gear noise.

Traditionally, gear designers have treated gear noise as an afterthought, using stress analysis for the basic gear design, and then learning after the gears are in operation that noise is a problem. It is hoped that some of the concepts presented here will encourage designers to do more to include noise reduction at the gear design stage.

## Cross-References

- [Friction in Gears](#)
- [Gear EHL Film Thickness and Wear Risk Analysis](#)
- [Gear Sliding](#)
- [Gear Tooth Modifications](#)
- [Helical Gears](#)
- [Involute Gear Profiles](#)
- [Non-involute Gear Profiles](#)
- [Spur Gears](#)

## References

- J. Börner, D.R. Houser, Influence of friction and bending moments on gear noise excitations, in *SAE Noise and Vibration Conference and Exhibition Paper 961816*, SAE Trans. **105**(6), 1669–1676 (1996)
- B. Hansen, M. Salerno, L. Winkelmann, Isotropic superfinishing of S-76C+ main transmission gears, in *Proceedings of Fall Technical Meeting, American Gear Manufacturers Assn*, Orlando, 2006, Paper 06FTM02.
- D.R. Houser, Gear noise and vibration prediction and control methods, in *Handbook of Noise and Vibration Control*, ed. by M.J. Crocker (Wiley, New York, 2007), Chap. 69
- D.R. Houser, F.B. Oswald, M.J. Valco, R.J. Drago, J.W. Lenski, Comparison of transmission error predictions with noise measurements for several spur and helical gears, in *Proceedings of 30th AIAA/ASME/SAE/ASEE Joint Propulsion Conference*, Indianapolis, 1994
- D.R. Houser, J. Harianto, The effect of micro-geometry and load on helical gear noise excitations, in *Proceedings of SAE Noise and Vibration Conference and Exhibition*, Traverse City, 2005, Paper No. 2005-01-2295
- D.R. Houser, J. Harianto, *Load Distribution Program Reference Manual*. Available from: GearLab – The Ohio State University, 2009

- N. Ozguven, D.R. Houser, Mathematical models used in gear dynamics – a review. *J. Sound Vib.* **121**(3), 383–411 (1988)
- R. Singh, D.R. Houser, Basic gear noise short course notes. Available from the GearLab, The Ohio State University, 2012
- R. Singh, Gear noise: anatomy, prediction and solutions, in *Proceedings, Inter-Noise 2009*, Ottawa, 2009
- D.B. Welbourn, Fundamental knowledge of gear noise – a survey, in *Proceedings of Noise and Vibration of Engines and Transmissions, Institute of Mechanical Engineering Automotive Division*, 1979, pp. 9–14

---

## Gear Noise Identification

- [Gear Dynamics](#)

---

## Gear Oil

- [Gear Lubricants](#)

---

## Gear Rattle

- [Gear Noise](#)

---

## Gear Shaping

- [Gear Manufacturing](#)

---

## Gear Sliding

ROBERT ERRICHELLO  
GEARTECH, Townsend, MT, USA

## Synonyms

[Entraining velocity](#); [Rolling velocity](#); [Slide/roll ratio](#); [Sliding velocity](#); [Specific sliding \(slide/sweep\) ratio](#)

## Definition

Sliding velocity is the relative velocity in a transverse plane of a common contact point between mating gear teeth.

As shown in Fig. 1, it is the vectorial difference between the two rolling velocities that are tangential to the tooth profiles and perpendicular to the line of action. See references AGMA (2005) and ISO (2007) for gear nomenclature and definitions of terms.

## Scientific Fundamentals

Gear teeth both roll and slide on one another. Figure 1 shows that the rolling velocity of the pinion,  $v_{r1}$ , and the rolling velocity of the gear,  $v_{r2}$ , linearly increase from zero at the interference points to a maximum at each end of the path (SAP and EAP) of contact. The distance between the  $v_{r1}$  and  $v_{r2}$  lines represents the sliding velocity. The amount of sliding is proportional to the distance from the pitch point, P, and is zero when the gear teeth contact at the pitch point and largest at the ends of the path. Sliding is oppositely directed on either side of the pitch point.

Figure 2 shows the directions of the rolling (R) and sliding (S) velocities on the driving and driven gear teeth. Contact on the driver tooth starts near the root of the tooth, rolls up the tooth, and ends at the tooth tip. Sliding is away from the driving gear pitchline. Contact on the driven tooth starts at the tooth tip, rolls down the tooth, and ends near the tooth root. Sliding is towards the driven gear pitchline.

Hertzian fatigue cracks, either macropitting or micropitting, that start at the gear tooth surface grow at a shallow angle to the surface and opposite to the slide directions. Consequently, as shown in Fig. 2, the cracks converge near the pitch line of the driver and diverge near the pitch line of the driven gear.

**Kinematics.** The following equations give velocity parameters that measure meshing characteristics of gear teeth. See reference AGMA (2003) for information on the variables used in (1)–(5).

## Rolling Velocities

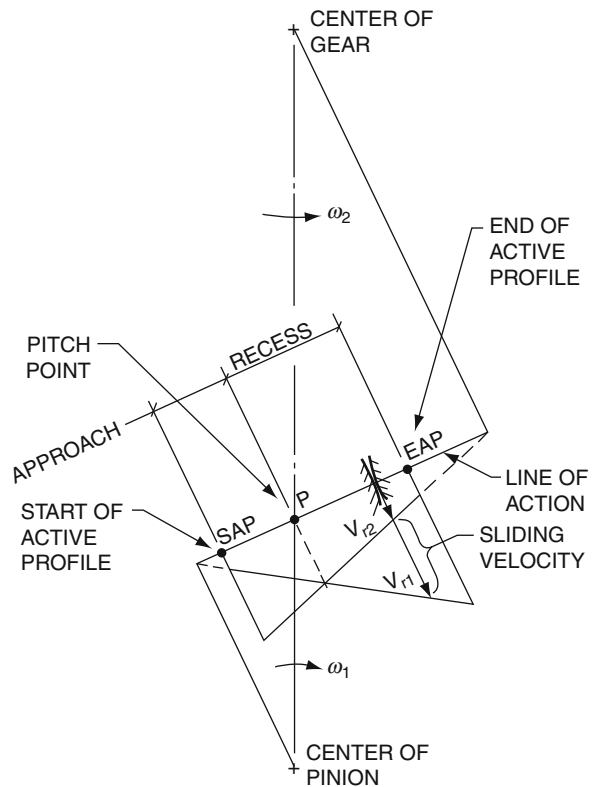
Rolling velocity is beneficial because it entrains lubricant between contacting teeth, increases oil film thickness, and reduces the number and severity of asperity contacts.

$$v_{r1} = \omega_1 * \rho_1 \quad (1)$$

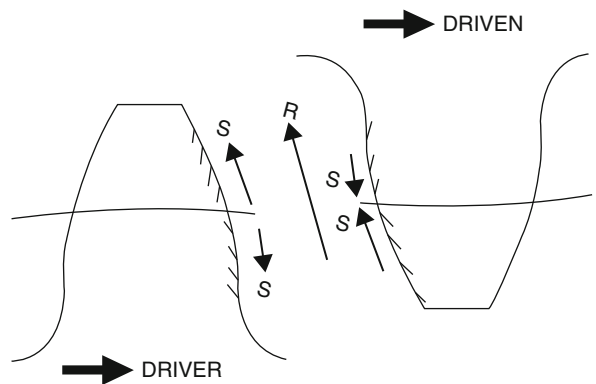
$$v_{r2} = \omega_2 * \rho_2 \quad (2)$$

Where:

$v_{r1}, v_{r2}$	Rolling velocities (pinion, gear), mm/s
$\omega_1, \omega_2$	Angular velocities (pinion, gear), rad/s
$\rho_1, \rho_2$	Transverse profile radii (pinion, gear), mm



Gear Sliding, Fig. 1 Rolling and sliding velocity



Gear Sliding, Fig. 2 Directions of rolling and sliding velocities

## Sliding Velocities

Sliding velocity is detrimental because it generates heat, reduces efficiency, and increases asperity distress. High sliding velocity generates frictional power loss by shearing the oil film and shearing boundary films on contacting



asperities. High sliding velocity also increases flash temperature and risk of scuffing.

$$v_{s1} = v_{r1} - v_{r2} \quad (3)$$

$$v_{s2} = -v_{s1} \quad (4)$$

## Entraining Velocity

$$v_e = v_{r1} + v_{r2} \quad (5)$$

Entraining velocity draws oil between contacting gear teeth and increases EHL film thickness.

## Specific Sliding Ratios

$$v_{ss1} = \frac{v_{s1}}{v_{r1}} \quad (6)$$

$$v_{ss2} = \frac{v_{s2}}{v_{r2}} \quad (7)$$

Specific sliding ratio is a contact severity parameter that is:

- A dimensionless ratio dependent on tooth geometry and independent of pitch line velocity.
- Positive in addenda, ranging from zero at the pitch point to +1 at the interference point of the mating gear.

- Negative in dedenda, ranging from zero at the pitch point to  $-\infty$  at the interference point.

Generally,  $|v_{ss1}| < 1.0$  is recommended, and  $|v_{ss1}| \leq 0.5$  is preferred, for good resistance to macropitting, micropitting, and scuffing. Specific sliding ratio is also called slide/sweep ratio.

## Slide/roll Ratio

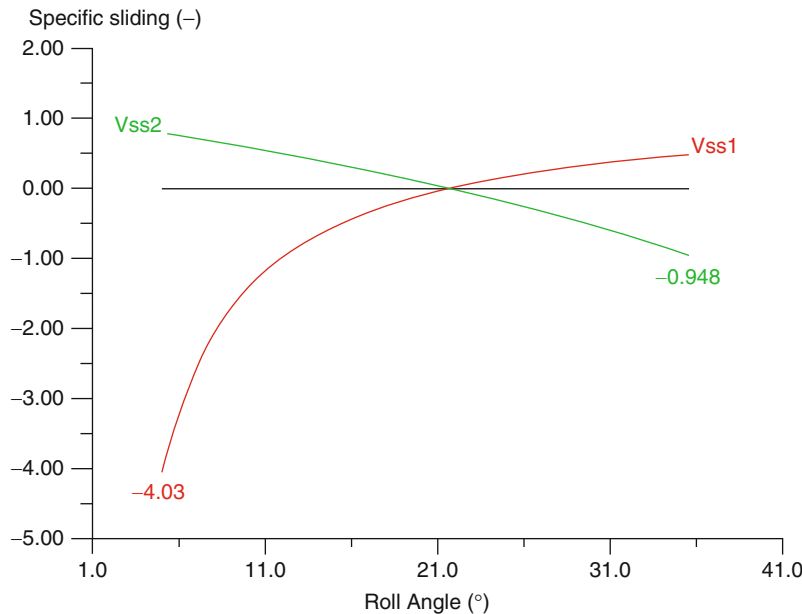
$$v_{se} = \frac{v_{s1}}{v_e} \quad (8)$$

Some literature uses mean entraining velocity, in which case the slide/roll ratio is:

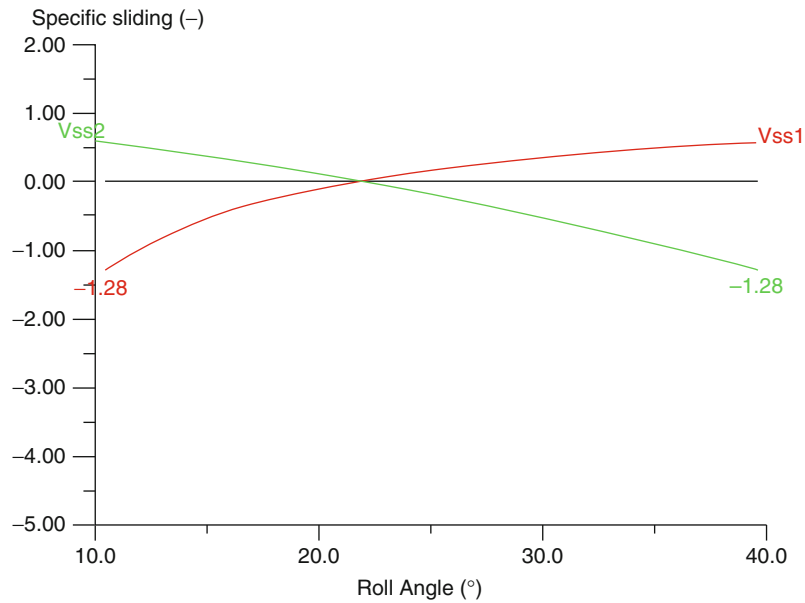
$$\bar{v}_{se} = \frac{2v_{s1}}{v_e} \quad (9)$$

**Negative Sliding.** Gear teeth dedenda have negative sliding (direction of rolling velocity is opposite sliding velocity). Negative sliding is significant because it promotes Hertzian fatigue. It is believed that negative sliding allows oil to enter surface cracks and accelerate crack growth by the hydraulic pressure propagation mechanism first proposed by Stewart Way (1935).

**Gear Design.** Profile shift (AGMA 1998) can be used to balance and minimize the specific sliding ratio that occurs at the ends of the contact path. Figure 3 pertains to a helical gearset that has high specific sliding ratio  $v_{ss1} = -4.03$  that



Gear Sliding, Fig. 3 Gearset with high specific sliding ratio



**Gear Sliding, Fig. 4** Gearset with balanced specific sliding ratio

occurs at the SAP of the pinion. Figure 4 pertains to a similar gearset, except the profiles of the teeth have been shifted outward on the pinion and inward on the gear in order to reduce and balance the specific sliding ratio to  $v_{ss1} = v_{ss2} = -1.28$ .

## Key Applications

Rolling velocity, sliding velocity, and entraining velocity are used in calculations of gear tooth EHL film thickness and contact temperature. Specific sliding ratio is a useful parameter for optimizing gear geometry. Industrial standards such as (AGMA 2003) provide fundamental rating formulas to evaluate gear tooth load capacity.

## Cross-References

► [Sliding Wear](#)

## References

- AGMA 913-A98, *Method for Specifying the Geometry of Spur and Helical Gears*, AGMA (1998)
- AGMA 925-A03, *AGMA Information Sheet – Effect of Lubrication on Gear Surface Distress*, AGMA (2003)
- ANSI/AGMA 1012-G05, *Gear Nomenclature, Definition of Terms with Symbols*, AGMA (2005)
- ISO 21771:2007(E), *Gears – Cylindrical Involute Gears and Gear Pairs – Concepts and Geometry*, ISO (2007)
- S. Way, Pitting due to rolling contact. *ASME J. Appl. Mech.* **57**, A49–A114 (1935)

## Gear Stamping

► [Gear Manufacturing](#)

## Gear Surface Pitting Failure and Pitting Life Analysis

ROBERT ERRICHELLO

GEARTECH, Townsend, MT, USA

## Synonyms

Contact fatigue of gear surfaces; Hertzian contact fatigue; Macropitting; Macropitting resistance; Pitting failure; Pitting in gear surfaces; Rolling contact fatigue; Surface distress; Surface durability.

## Definition

Hertzian fatigue is caused by repeated contact stresses that cause surface or subsurface fatigue cracks and detachment of material fragments from gear tooth surfaces. ANSI/AGMA 1010-E95 (1995) describes the characteristics of the failure



**Gear Surface Pitting Failure and Pitting Life Analysis,**  
Fig. 1 Macropitting on a helical pinion

mode. AGMA 925-A03 (2003a) gives information on effects of lubrication on gear surface distress.

Figure 1 shows macropitting on a helical pinion.

### Scientific Fundamentals

Gear tooth load capacity based on Hertzian fatigue resistance is calculated according to industrial standards such as ANSI/AGMA 2101-D04 (2004) and ISO 6336-2:1999 (1999). Although Hertzian stress is a major factor governing Hertzian fatigue, the properties of the lubricant and its additives exert a strong influence (Errichello 1992). Therefore, to help prevent Hertzian fatigue, the physical and chemical properties of the lubricant should conform to the requirements of ANSI/AGMA 9005-D94 (1994a). It is especially important that lubricant viscosity be appropriate for the pitchline velocity of the gears to ensure adequate Hertzian fatigue resistance.

### Hertzian Stress Analysis

Industrial standards (AGMA 2004; ISO 1999) calculate Hertzian stress with an equation similar to (1). See references (AGMA 2003a, 2004) for information on the variables used in (1)–(4).

$$\sigma_H = \frac{2w_n}{\pi b_H} \quad (1)$$

$$w_n = \frac{F_{wn} K_O K_H K_V}{L_{min}} \quad (2)$$

$$b_H = \left( \frac{8X_\Gamma w_n \rho_n}{\pi E_r} \right)^{1/2} \quad (3)$$

$$E_r = 2 \left( \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2} \right)^{-1} \quad (4)$$

where

$\sigma_H$	Hertzian contact stress, N/mm <sup>2</sup>
$w_n$	Normal unit load, N/mm
$b_H$	Semi-width of the Hertzian contact band, mm
$F_{wn}$	Total load normal to the gear tooth profiles at the operating pitch point, N
$K_O$	Overload factor accounts for externally applied loads induced by dynamics of the gear system, dimensionless
$K_H$	Load distribution factor accounts for non-uniform distribution of load along the lines of contact between mating gear teeth, dimensionless
$K_V$	Dynamic factor accounts for internally generated gear tooth loads induced by non-uniform meshing action of gear teeth, dimensionless
$L_{min}$	Minimum contact length, mm
$X_\Gamma$	Load sharing factor, dimensionless
$\rho_n$	Normal relative radius of curvature, mm
$E_r$	Reduced modulus of elasticity, N/mm <sup>2</sup>
$\nu_1, \nu_2$	Poisson's ratio (pinion, gear), dimensionless
$E_1, E_2$	Modulus of elasticity (pinion, gear), N/mm <sup>2</sup>

### Pitting Life Analysis

Industrial standards (AGMA 2004; ISO 1999) use empirical methods to relate calculated values of Hertzian stress to empirical values of allowable contact stress with an equation similar to (5). See reference (AGMA 2004) for information on the variables used in (5).

$$S_H = \frac{\sigma_{HP} Z_N Z_W}{\sigma_H Y_\theta Y_Z} \quad (5)$$

where

$S_H$	Safety factor accounts for uncertainties or statistical variations in design analysis, material properties, and manufacturing tolerances, dimensionless
$\sigma_{HP}$	Allowable contact stress for a baseline design life such as $10^7$ cycles, N/mm <sup>2</sup>
$Z_N$	Stress-cycle life factor adjusts the allowable stress for a design life different than the baseline design life, dimensionless
$Z_W$	Hardness ratio factor accounts for work hardening that increases gear capacity, dimensionless
$\sigma_H$	Hertzian contact stress, N/mm <sup>2</sup>
$Y_\theta$	Temperature factor accounts for reduction of fatigue strength at high temperature, dimensionless
$Y_Z$	Reliability factor accounts for statistical variation in fatigue strength, dimensionless

## Key Applications

Industrial standards such as (AGMA 2004; ISO 1999) provide fundamental rating formulas to evaluate gear tooth load capacity. Application standards tailor the fundamental formulas to particular applications such as wind turbine gears (AGMA 2003b) or marine gears (AGMA 1994b).

## Cross-References

- Gear Lubricants
- Gear Lubrication
- Gear Materials
- Rolling Contact Fatigue

## References

- AGMA, ANSI/AGMA 9005-D94, *Industrial Gear Lubrication*, (1994a)
- AGMA, ANSI/AGMA 6032-A94, *AGMA Standard – Standard for Marine Gear Units: Rating*, (1994b)
- AGMA, ANSI/AGMA 1010-E95, *AGMA Standard – Appearance of Gear Teeth – Terminology of Wear and Failure*, (1995)
- AGMA, AGMA 925-A03, *AGMA Information Sheet – Effect of Lubrication on Gear Surface Distress*, (2003a)
- AGMA, ANSI/AGMA/AWEA 6006-A03, *AGMA Standard – Standard for Design and Specification of Gearboxes for Wind Turbines*, (2003b)
- AGMA, ANSI/AGMA 2101-D04, *AGMA Standard – Fundamental Rating Factors and Calculation Methods for Involute Spur and Helical Gear Teeth*, (2004)
- R. Errichello, Friction, Lubrication, and Wear of Gears, in *ASM Handbook*, vol. 18 (ASM International, Metals Park, 1992), pp. 535–545
- ISO, ISO 6336-2:1999, *ISO International Standard – Calculation of Load Capacity of Spur and Helical Gears – Part 2: Calculation of Surface Durability (Pitting)*, (1999)

## Gear Surface Treatment

CHARLES D. SCHULTZ

Beyta Gear Service, Winfield, IL, USA

## Synonyms

Coating; Honing; Lapping; Shaving; Superfinishing

## Definition

Gear surface treatment is any process performed on the operating surfaces of gear teeth to improve their performance.

## Scientific Fundamentals

Manufacturing processes produce gear tooth surfaces with characteristics that may not provide optimum operating performance. In some cases, these characteristics are

considered “errors” in that the processes used are unable to produce the geometrical shape or accuracy required by the drawings. In other cases, the parts comply with the drawings but additional processing is needed to improve operating performance. In the earliest days of gear making, all gears were “hand fitted” to overcome inconsistencies in their manufacture. The introduction of machine-cut teeth initially reduced the need for “fitting,” but as loads and rotational speeds increased, gear makers began to “adjust” contact patterns again to reduce noise and vibration, improve load capacity, and extend service life.

A variety of “surface treatment” options are employed today. All but one involve removing material from the tooth surface, thus reducing the “peak to valley” variation and improving the effective film thickness of the lubricant. The final technique applies a substance to the surface, filling in some of the valleys, and providing at least a temporary improvement to lubrication conditions. Each method, described in more detail below, has its place in the gear trade.

## Surface Finish Versus Manufacturing Method

AGMA 2000 does not include surface finish in its tolerancing. Table 1 shows the AGMA quality levels normally produced by common production methods. Comparing this table with Table 2 on Surface Finish vs. Production Method shows that there is an indirect relationship between “quality” and “surface finish.” Specifying a tooth surface finish may require costly gear finishing processes. Surface finish has an effect on lubricant specific film thickness. While no consensus “standard” has been published on which lubricant viscosities are needed with which surface finishes, it is clear that higher viscosity oil is needed when coarser finishes are present. The use of the higher viscosity lubricant may or may not be possible in some applications due to cold starting conditions, thermal considerations, or other issues.

## Methods for Improving Surface Finish

### Lapping

Lapping – the oldest gear “finishing” method – involves running a set of gears with an abrasive fluid in place of the lubricant. This process was developed to adjust for cutting inaccuracies, increase backlash, and improve surface finish. Lapping is no longer widely used as sophisticated gear inspection techniques have revealed that excessive lapping can destroy the involute form cut into the teeth. Modern gear cutting equipment can usually produce parts that do

**Gear Surface Treatment, Table 1** Achievable AGMA quality levels by manufacturing method

Manufacturing method	Runout	Involute profile	Lead	Spacing (Pitch)	Relative cost
Hobbing (Class B hob)	Q8 -10	Q8 – 9	Q8 – 9	Q8 – 9	1.00–1.25
Hobbing (Class A hob)	Q9 -11	Q8 – 9	Q9 – 11	Q8 – 10	1.25–1.50
Hobbing (Class AA hob)	Q9 -12	Q8 – 11	Q9 – 11	Q9 – 11	1.50–1.75
Shaping (Commercial cutter)	Q8 -10	Q8 – 10	Q8 – 11	Q8 – 10	1.25–1.50
Shaping (Precision cutter)	Q9 -11	Q9 – 10	Q9 – 11	Q9 – 11	1.50–1.75
Shaving	Q10 -12	Q8 – 10	Q8 – 12	Q8 – 12	2.00–2.50
Grinding	Q9 -14	Q9 – 14	Q8 – 14	Q9 – 14	3.00–4.00

**Notes**

1. Lower quality levels are generally achievable under most conditions
2. Upper quality levels require special controls on blanks, tooling, and machinery. This can increase costs significantly
3. Relative costs compared with Class B hobbing for operations needed to finish gear teeth only. Material and heat treat costs are not included in this comparison
4. If heat treating is done after tooth finishing, quality level can drop by two levels or more

not require lapping and lapping requirements noted on older drawings are frequently ignored. The most common use of lapping today is as a “last resort” in solving field problems or making very fine adjustments in backlash for pump or instrument gears. In some cases, gears are lapped with “dummies” (made of cast iron or another soft material) to replicate the mating part. This reduces the tendency to damage the involute form.

**Shaving**

Shaving is a gear “finishing” method that is used to improve surface finish and gear geometry (lead and involute). A serrated gear-like cutter is rotated and fed axially while in mesh with the part. The part teeth must be cut with a special “pre-shave” hob or shaper cutter that finishes the root area of the part but leaves “stock” on the sides of the teeth. The cutter axis is at an angle to the part axis that creates a shaving or planing action as the cutter moves axially across the gear face. Spur, helical, and double helical (with gap) gears can be shaved. Both internal and external teeth can be accommodated if proper tooling can be designed. Shaving is most successful on parts less than 50 HRC.

**Honing**

The honing process is similar to shaving except that the cutters are coated with an abrasive material. As shaving machines can also be adapted to honing, most of the limitations discussed above still apply. Honing has been used to produce very finely polished surface finishes (as low as 6AA) and to finish surface hardened gears that are not suitable for gear grinding.

**Gear Surface Treatment, Table 2** Achievable tooth flank surface finishes (AA) in micro-inches by manufacturing method

Tooth Size	Effort	Hobbing	Shaping	Shaving	Grinding
1–3 DP	Normal	125	80–125	63	32
	Extra	80	80	32	16
3–10 DP	Normal	80	63–80	63	32
	Extra	63	63	32	16
10–24 DP	Normal	80	63–80	50–32	32
	Extra	63	63	16	16
24–40 DP	Normal	80	63	50–32	32
	Extra	63	63–32	16	16
40 DP & up	Normal	80	63	Not	32
	Extra	63	63–32	Practical	16

**Notes**

1. Normal effort involves typical production feeds and speeds
2. Extra effort involves special controls and procedures on tools and machines. Cycle times
3. Finishes shown are for through hardened steel of 230–310 HB.
4. Finish may be poorer on steel below 230 HB and above 310 HB
5. Surface finish may be slightly better in brass, bronze, aluminum, or stainless steel provided proper feeds and speeds are selected
6. Surface finish for surface hardened gears that are not finished after heat treating may be slightly worse due to scale removal operations

**Superfinishing**

Superfinishing is the generic term for improving the surface finish with a combination of mechanical polishing and chemical etching. Chemically accelerated vibratory

finishing levels the “peaks” on the part surface. With proper process controls, finishes as fine as 4  $\mu$ -in. AA can be achieved. As with lapping and polishing, the process must be carefully monitored to avoid damage to the involute form.

## Coating

Coating is a process of applying a substance to the tooth surfaces. Depending on the coating composition, a combination of temperature and pressure is needed to insure proper adhesion to the tooth surface. This may prevent the processing of carburized or induction hardened gears as exposure to temperatures over 300 °F may reduce the surface hardness below acceptable levels. Nitrided parts must have the “white layer” removed prior to coating to insure proper adhesion.

Coatings can provide decorative or rust preventative properties in addition to improving surface roughness measurements. Soft coatings such as manganese phosphate, copper, and silver provide a sacrificial protection against micropitting and scuffing during start-up or run-in periods. Thin, super-hard coatings such as diamond-like carbon or boron carbide increase resistance to abrasive wear, scuffing, and micropitting.

## Key Applications

As noted in Table 2, surface finish is a complex function of material, heat treatment, tool geometry, depth of cut, lubricant/coolant, and cutting speed. Generally speaking, the larger the chip, the worse the surface finish. Within a particular process, adjustments to these factors can affect the resulting finish within a relatively narrow range. Some portions of a gear tooth, such as the root fillet, may have significantly worse finish than others. This may be a result of the severity of the cutting action in the local area or damage due to chip entrapment. Achieving finishes better than those listed in Table 2 may not be economically feasible, and alternate finishing methods such as lapping, polishing, superfinishing, or coating may be needed.

## Cross-References

- Gear Cutting Tools
- Gear Manufacturing
- Gear Manufacturing Machines

## References

- D.W. Dudley, *The Evolution of the Gear Art*. American Gear Manufacturers Association Library of Congress Catalog Card Number: 72-78509
- D.W. Dudley, *Practical Gear Design*. McGraw-Hill Library of Congress Catalog Card Number: 53-11476

D.W. Dudley, *Gear Handbook*. McGraw-Hill Library of Congress Catalog Card Number: 61-7304

C.D. Schultz, *An Introduction to Gear Design*. Privately published 1987–1997. Available for free download at [www.beytgear.com](http://www.beytgear.com)

L. Winkelmann, et al., The capacity of superfinished vehicle components to increase fuel economy Part 1. *Gear Technol.*, 26(1), Jan/Feb 2009

## Gear Tooth Modifications

JOHN R. COLBOURNE

Department of Mechanical Engineering, University of Alberta, Edmonton, AB, Canada

## Definition

Profile modification is the term used to describe small changes in a gear tooth profile from the theoretical conjugate profile, to compensate for manufacturing errors and for changes in the tooth shape under load.

## Scientific Fundamentals

### Introduction

The purpose of profile modification is to try to ensure smooth meshing, even though there may be small manufacturing errors, and the teeth may deflect under load. In theory, the teeth should be designed so that under load the deformed profiles should be conjugate. This is only theoretically possible under one specified load, usually the design load, so it is recognized that under zero or light load the teeth will not be exactly conjugate.

There are three principal types of profile modification, which will be described here. The first type is tip and root relief, the second is lead modification, and the third is crowning.

### Tip and Root Relief

When two loaded gears rotate, the meshing teeth will deflect a small amount, causing the driven gear to lag slightly behind its theoretical position. As the contact point on one tooth pair reaches the highest point of single-tooth contact on the driving gear, a second tooth pair should come smoothly into contact at the limit diameter of the driving gear. Due to the lag of the driven gear, the tooth tips of the driven gear will dig into the flanks of the driving gear, near the limit diameter, and this will cause vibration and eventually damage to the teeth.

In order to avoid this problem, it is common practice to reduce the tooth thickness of the gears near the tips and



the limit diameters, and this modification of the profiles is known as tip and root relief. However, since root relief on one gear has the same effect as tip relief on the meshing gear, the most common procedure is to provide tip relief on both the driving and the driven gears, and no root relief.

The amount of tip relief is generally specified in the transverse section, and is the distance of the tooth tip from the theoretical conjugate profile, measured normal to the conjugate profile. It is also necessary to specify the point of the profile at which the tip relief starts, and this is generally near the addendum mid-point, or slightly closer to the tip. It is important that the transition from the conjugate profile to the tip relieved segment should be as smooth as possible, so the deviation from the conjugate profile to the actual tooth profile should be a second degree or higher degree function.

A design procedure for specifying TR, the amount of tip relief, is given by the AGMA (AGMA 6002-B93 1993):

$$F_t = T/R_p \quad (1)$$

$$w = F_t/b \quad (2)$$

$$TR = 3.0 w/10^7 \quad (3)$$

where  $F_t$  is the tangential load,  $T$  is the torque,  $R_p$  is the pitch circle radius,  $b$  is the face-width, and  $w$  is the unit load, all measured in Imperial units.

In the metric system of units, the torque is generally expressed in Nm, and the lengths in mm. The equivalent equations are then

$$F_t = 1000 T/R_p \quad (4)$$

$$w = F_t/b \quad (5)$$

$$TR = 4.351 w/10^5 \quad (6)$$

The AGMA value applies to both gears of a gearset. However, it is standard practice to design the pinion of a gear pair with a greater tooth thickness than the gear, in order to balance the root fillet bending stresses. When this is done, the pinion is probably stiffer in bending than the gear, so the pinion tip relief can be less than that of the gear. An alternative theory has been suggested by Dudley (Dudley 1991), who argued that when there is no tip relief the main damage occurs at the start of contact, so the tip relief on the driven gear should be greater than that on the driving gear. In place of the value 3.0 in (3), he suggested the value 3.5 for the driven gear, and 2.0 for the driving gear, for a pair of spur gears, and stated that these values might need to be modified by experience.

All the tip relief values given by the AGMA and by Dudley are for involute gears. In the case of non-involute gears, it is necessary to compare their bending stiffness with that of involute gears, and to modify the tip relief values accordingly. This is best done using finite element analysis to calculate the bending stiffness of the non-involute teeth. Even in the case of involute gears, there are commercial programs available that will calculate the bending stiffness of the teeth, and then provide values for the tip relief.

### Lead Modification

A typical gear is mounted on a shaft, which is supported at each end by a bearing. One end of the shaft is free to rotate, while the other end has a torque  $T_1$  applied to it, which is balanced by the tooth forces in the gear. Since the tooth forces are spread evenly over the face width of the gear, at least in theory, the torque in the gear will vary in a linear manner from zero at one end to  $T_1$  at the other. This torque will cause a slight twist of the gear, over and above the tooth rotation corresponding to the helix angle. If no correction is made for this twisting, the tooth forces will be concentrated at the end of the gear where the torque is applied. The purpose of lead modification is to ensure that the tooth forces under full load are spread fairly evenly over the face width.

The amount of twisting is greater in the pinion than in the gear, since the pinion root diameter is smaller than that of the gear. For the purpose of calculating the amount of twist, the pinion will be treated as if it is a shaft of diameter equal to its root diameter  $D_{\text{root1}}$ .

If the axial  $z$  coordinate origin is at the end face where the torque is zero, and the face-width is  $b$ , then the torque  $T$  at point  $z$  on the pinion axis is given by

$$T = z T_1/b \quad (7)$$

The twist at the transverse plane  $z$  on the gear can be represented by the function  $\phi(z)$ , and then the derivative of  $\phi$  with respect to  $z$  is given by

$$d\phi/dz = T/GJ_1 \quad (8)$$

where  $G$  is the shear modulus, and  $J_1$  is the polar moment of inertia of the assumed shaft, given by

$$J_1 = \pi D_{\text{root1}}^4/32 \quad (9)$$

Combining (7) and (8) and integrating, we obtain an expression for the total twist  $\phi_1$  of the pinion at the end face  $z = b$ :

$$\phi_1 = T_1 b/2GJ_1 \quad (10)$$

The rotation  $\Delta\theta$  of the tooth profile between one end face and the other, due to the helix angle  $\beta$ , is

$$\Delta\theta = 2\pi b/L_1 \quad (11)$$

where  $L_1$  is the lead of the pinion.

Hence the total rotation  $\Delta\theta'$  of the tooth profile, due to the helix angle and the twisting, is given by

$$\Delta\theta' = \Delta\theta \pm \varphi_1 \quad (12)$$

where the sign depends on whether the torque is applied in a direction to increase or decrease the helix angle. The pinion will effectively have a helix angle  $\beta'$  given by

$$\tan \beta' = \tan \beta (1 \pm L_1 \varphi_1 / 2\pi b_1) \quad (13)$$

and for correct meshing the gear must have the same helix angle. The procedure by which the gear helix angle is altered from  $\beta$  to  $\beta'$  is called lead modification.

The analysis presented here is only approximate. The situation is similar to that described in the discussion on tip relief, in that more accurate values for the lead modification can be found by using finite element analysis to determine the torsional flexibility of the pinion.

### Crowning

In addition to torsional twist, the pinion body bends, and it is often necessary to compensate for the bending deflection by crowning the teeth. Gear teeth are described as crowned if the tooth thickness is slightly reduced near the end faces. This is done to prevent heavy contact at the end faces, which can cause damage and even breakage.

There is not much material publicly available to determine the optimum amount of crowning. Dudley (1991) has suggested that the reduction in tooth thickness will generally be between 0.0005 and 0.002 in., but the correct amount for any particular gearset will clearly depend on the module of the teeth, and the stiffness of the supporting shafts and bearings. Determination of an accurate value for the crowning would require a full analysis of the shafts and bearings.

### Lead Modification and Crowning

To obtain uniform load intensity across the face width, it is common practice to alter the helix angle as a slope correction and add a crown. This combined modification compensates for both torsional twist and bending deflection.

### Key Applications

Gears whose tooth profiles are initially conjugate will undergo deformations under heavy load, so that they

cease to be conjugate. The tooth profile modifications are designed so that the deformed teeth at rated loads are as close as possible to being conjugate.

### References

- AGMA 6002-B93, *Design Guide for Vehicle Spur and Helical Gears* (American Gear Manufacturers Association, Alexandria, 1993)
- D.W. Dudley, *Dudley's Gear Handbook* (McGraw-Hill, New York, 1991). Chap. 4

## Gear Vibration

- [Gear Dynamics](#)

## Gear Vibration and Noise

- [Gear Noise](#)

## Gear Whine

- [Gear Noise](#)

## Gearing Energy Losses

- [Gear Efficiency](#)
- [Gear Lubrication](#)

## Gearing Friction

- [Gear Efficiency](#)

## Gears and Mixed Lubrication

- [Mixed EHL in Gears](#)

## Gecko Toe Surface

YU TIAN

State Key Laboratory of Tribology, Tsinghua University,  
Beijing, People's Republic of China

### Synonyms

GS-Gecko Surface

### Definition

Gecko surface includes the hair pad of living geckos and the bionic surfaces with micro structures with strong adhesion and ► [friction](#) forces.

### Scientific Fundamentals

#### Structure of Gecko Toe Pad

Geckos have extraordinary capabilities of climbing vertical walls and upside-down ceilings with strong adhesion strength and easy removal. Their toes can stick to almost any surface through ► [van der Waals forces](#), whether hydrophilic or hydrophobic, rough or smooth, and under nearly any condition (even in a vacuum). The gecko's mechanisms of moving and adhering have attracted very wide interest both in the science and engineering communities. Not until a 100 years ago, using optical microscopy, was the adhering ascribed to the hair-like microstructure of gecko toe pads. About a half century ago, using by electron microscopy, the fine structure of hairs was discovered to be a multiple hierarchical split ends and spatulae nanostructure at the tip of each hair. The mechanism of dry adhesion of gecko hairs with hierarchical structures down to the nanometer scale and rapid switching between attachment and detachment has been the focus of scientific study for more than a century.

The extraordinary climbing ability of the gecko is based on the hierarchical structure of its toes, which contain setal arrays consisting of hundreds of spatulae on each seta. The macro-, meso-, micro-, and nanoscale structures of the hierarchical structure are shown in [Fig. 1](#). There are five toes on each foot, and about 20 rows of sticky setal arrays on a gecko toe, with thousands of setae. One Tokay gecko foot has approximately 14,400 setae per mm<sup>2</sup>. Each seta terminates with 100–1,000 spatulae on its end. A single seta is about 110 μm in length and 4.2 μm in diameter. A single spatula is thin and has a roughly triangular fan shape that is approximately 0.2 μm in length and 10 nm in thickness. The huge number of spatulae promises a large “real” contact area between the spatulae and the

adhered surface to provide high adhesion and friction forces via the weak but universal van der Waals force. It has been found that more than a thousand gecko species have adhesive toe pads and some of them also have setae on their tails (Autumn [2006](#)).

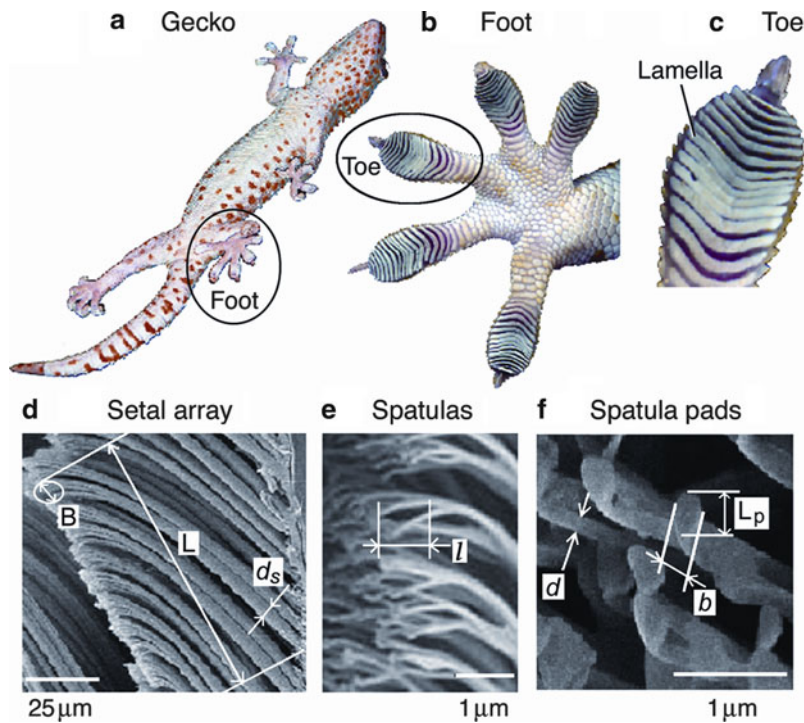
#### Adhesive Force at Different Structure Levels

Due to the size of setae and spatulae at the micro/nanometer scale, it is hard to manipulate and measure the adhesive forces of a single seta or a single spatula. However, there are still some successful measurements on the adhesion or friction forces at different structure levels. The two front feet of a tokay gecko, with a pad area of about 200 mm<sup>2</sup>, can withstand about 20 N friction force on a surface. The friction force of an isolated setal array is about 0.37 N with area of 0.99 mm<sup>2</sup>, versus the single toe with 4.3 N and an area of 0.19 cm<sup>2</sup>. The maximum friction force observed for a single seta (with about 100–1,000 spatulae) is about 200 μN, and the adhesive force is 20–40 μN. For an isolated seta, with most terminal branches of seta machined away to isolate a few single spatula and then glued perpendicular to the end of an AFM cantilever, the adhesion force of a single spatula was measured to be about 10 nN. However, single setae tests have been found to be either much less or more sticky than predicted in whole-animal measurements (Autumn and Gravish [2008](#)).

#### The Origin of Adhesion Force

While adhesion and friction forces of gecko feet have been tested at different structural levels, research has also concerned the origin of the adhesion force. The most widely proposed forces are the van der Waals force between two solid surfaces and the capillary adhesion force of a liquid film confined between two solid surfaces. Since the discovery of the microstructures of setal tips by Ruibal and Ernst in 1965, real contact area is considered to be increased by the spatulae and the strong adhesion force of gecko feet comes from the molecular interactions between surfaces rather than mechanical interlocking. Subsequent experimental research on gecko adhesion showed that the surface energy of the substrate, rather than its texture, greatly affected the adhesion strength.

However, because the capillary adhesion force is the principal adhesion force in many insects, frogs, and the marsupial feathertail glider, researchers have wondered whether it is similar to gecko adhesion, even though geckos usually lack glands on the surface of their feet. However, moisture in the air might provide a very thin liquid film between their spatulae and the adhered substrate. Layers of water molecules widely exist on



Gecko Toe Surface, Fig. 1 The hierarchical structures of gecko

hydrophilic surfaces at an ambient humidity and usually result in strong adhesion between surfaces. An experimental observation that geckos cannot adhere to a superhydrophobic polytetrafluoroethylene (PTFE, Teflon) would support the capillary hypothesis. In 2002, a carefully designed experiment for the adhesion force of a gecko foot to both hydrophobic and hydrophilic substrate in vacuum was done. For both substrates, the adhesion forces are strong but a bit higher for the hydrophilic surface. This experiment verified that gecko adhesion mainly depends on the van der Waals force. Work by Artz et al. showed that low humidity cannot form a continuous water layer, which might contribute to a significant capillary force, further supporting this conclusion. On the other hand, the adsorbed water layer on a substrate can increase the surface energy and enhance the van der Waals force between setae and substrate.

### The Continuum Contact Theory

For a Hertz contact, without considering the surface force, there is no adhesion force. Only when the surface force originating from the van der Waals forces or [surface forces](#) between surfaces comes into consideration of contact is there an adhesion force. The JKR, MD, and DMT contact models are the ones that consider the surface

forces in elastic sphere/plane contact. By simplifying the end of spatulae as cylinders terminated with hemispherical ends, the adhesion between spatulae and substrate can be calculated by JKR and other models. Using the forces measured in a single spatula, the calculated size of the JKR sphere is close to the real size of a spatula. Using values of  $R = 100 \text{ nm}$  and  $\gamma = 50 \text{ mJ/m}^2$ , where  $\gamma$  is the adhesion energy between the sphere and the surface, the predicted adhesion force for a gecko spatula is  $F = 23.6 \text{ nN}$ , which is twice the value measured by AFM by Huber et al. A simple further analysis by contact model shows that a higher level of hierarchical splitting of cylindrical structures results in a larger adhesion force. As a result, the shape of the terminating end is important. By analyzing two cases of contact scaling, self-similarity and curvature invariance, Artz et al. suggested that with much larger body mass, both the seta diameter and the radius of curvature have to be reduced. Furthermore, Spolenak et al. concluded that the spatula shape should have a very strong effect on adhesion force at sizes above  $100 \text{ nm}$  and particularly above  $1 \text{ }\mu\text{m}$ . In order to generate high adhesion forces, the optimum size and shape of artificial contact elements should match. However, the hemisphere approximate was not so accurate because of the various shapes in different biological contacts. Using the approach of Kendall,

spatulae may also be modeled as nanoscale adhesive tape. The Kendall peel model prediction is close to the actual dimension of the spatula (Gao and Yao 2004).

Since the microstructure of gecko seta consists of millions of high aspect ratio fibers, the effective elastic modulus is much lower than the Young's modulus of  $\beta$ -keratin, which is reported to be 2.5 GPa in bird feathers and 1.3–1.8 GPa in bird claws. It is well accepted that the main deformation of setae is bending, so the seta of gecko can be modeled as a cylindrical cantilever beam under a lateral load at its tip. It is easy to understand that the small fibers at the end of seta lead to a flexible deformation and that the setae hairs can follow various substrates. Although a tight contact can usually be obtained, the surface roughness still affects the real contact area and the resulting adhesion force. Huber et al. measured the adhesion of a single seta and substrates with different surface roughness with the help of AFM and showed the minimal adhesion for a root mean square roughness ranging from 100 to 300 nm (Persson and Gorb 2003).

### The Coupling of Adhesion and Friction Force

Because geckos require large adhesion and friction forces to move or climb, there exists a question of how geckos manage to detach their feet within 15 ms with a very small force. Experiments show that the friction forces of gecko setae are much higher than the adhesion forces, and there is a critical angle to release gecko toes and isolated seta. The relationship between adhesion and friction forces is one of the most fundamental issues for gecko attachment and detachment. During the detachment test of a single seta, Autumn et al. found that when the setal shaft makes an angle of  $30^\circ$  to the substrate, it is spontaneously detached with a very small adhesion force.

Autumn et al. thought that gecko adhesion depends directly on the shear force, i.e., the friction force in the gripping direction, and proposed a “frictional adhesion” model. By taking into account the final two levels of the hierarchical structures of setae and the macroscopic action of gecko toes, the research of how the friction force contributes to the adhesion force has been done. Based on the tape-like thin plate structure of spatulae and the van der Waals interaction between the spatulae and substrate, attachment and detachment of a gecko foot has been analyzed theoretically by Tian et al. (2006). Results show that the adhesion and friction forces originate from the van der Waals forces between the submicron-sized spatulae and the substrate, which are controlled by the macroscopic actions of the gecko toes. The pulling force of a spatula with a pulling angle between  $0^\circ$  and  $90^\circ$  has a normal adhesion force contribution, produced at the

spatula-substrate bifurcation zone, and a lateral friction force contribution from the part of spatula still in contact with the substrate. High net friction and adhesion forces on the whole gecko are obtained by rolling down and gripping the toes inward to realize small pulling angles among the large number of spatulae in contact with the substrate. To detach, the high adhesion friction is rapidly reduced to a very low value by relaxing and rolling the toes upward and backward, which, mediated by the lever function of the setal shaft, peels the spatulae off perpendicularly from the substrates. By these mechanisms, both the adhesion and friction forces of geckos can be changed over three orders of magnitude, allowing swift attachment and detachment during gecko running. The results have an implication for the fabrication of dry adhesives and robotic systems inspired by the gecko's locomotion.

In addition, by pressing setal arrays against a smooth polymer surface and then sliding various shearing distances, the coupling of adhesion and friction force of gecko setal arrays has been studied. During gripping and releasing action of the gecko toes, the relationships between adhesion and friction force are different due to the highly anisotropic nature of gecko hair structures. In the releasing direction, the frictional behavior follows Amonton's law while in the gripping direction both the adhesion and friction forces were highly nonlinear.

Gecko setae are also known as a self-cleaning adhesive. Tokay geckos can recover their cling ability after a few steps on clean glass after applying  $2.5\text{-}\mu\text{m}$  radius microspheres to their feet. Past research has shown that when a gecko's toes were immobilized and applied by hand to the surface of a glass force plate, after four simulated steps on a clean glass surface, the gecko removed contaminated particles on their setae to recover the adhesive function to support their body weight by a single toe. Tests show that isolated arrays of setae have the self-cleaning function that does not necessarily need the action of the gecko, which means the particles do not have strong adhesion force to the setae. Contact models show that self-cleaning depends on the energetic disequilibrium of the attractive forces acting on dirt particles between the substrate and the spatulae.

## Key Applications

### Synthetic Adhesives Mimicking Gecko Foot-Hair

In addition to the adhesion and friction mechanism investigations during gecko locomotion, researchers have tried to fabricate gecko surfaces to realize an easy and strong dry adhesive by various nanomanufacturing methods. The research on the mechanism of dry adhesion of gecko



hairs has provided some design criteria, such as contact splitting, the aspect ratio of pillars decided by the stiffness of the material, and the principle of anti-self-adhesion.

With the development of nanotechnology, dry adhesives inspired by gecko hairs have been widely fabricated by polymers or multiwalled carbon nanotubes. The most commonly used method is to pour liquid polymers into a template with micro/nano-size hole array patterns and then freeze them. Removing the template, surfaces with regular pillar arrays are obtained. With this method, surfaces have been created that mimic gecko surfaces with high adhesion forces (Sitti and Fearing 2003). Especially if coating a waterproof adhesive, such as a mussel protein layer, on pillars, these fabricated surfaces perform reversibly high adhesion under water. Dry adhesive with vertically aligned carbon nanotubes grown on silicon substrate shows an adhesion force of about  $100 \text{ N/cm}^2$ , much larger than that of a real gecko surface. However, with the change of pulling angle to the surface, the normal adhesion force has changed only slightly, thus the fabricated surface is still not easily released from the substrate. To date, most synthetic adhesives inspired by gecko hairs have failed to perfectly mimic the rapid switching between strong attachment and easy detachment. And fabricating such structures remains a challenge.

### Gecko Robots

A robot that can move on a vertical wall provides the possibility of automating tasks that are traditionally done manually, like cleaning tall buildings, performing inspections in dangerous environments such as poisonous gas environments, nuclear power plants, and war fields, and in space vehicles. It can afford great human safety and also better economic efficiency. By incorporating traditional robot structure design and control methods with a mimicked gecko surface, researchers are trying to develop high-performance robots to accomplish the tasks listed above. Much progress has been achieved, however, the bottleneck of this application remains the fabrication of a high-performance gecko surface.

### Other Possible applications

The adhesion and friction controllable gecko mimic surface can also be used in MEMS switching, wafer alignment and manipulation, micro-parts grabbing and transportation, and other micromanipulation. Because a dry adhesive gecko surface could be applied directly to any surface, it is conceivable that it might replace screws, glues, and interlocking pieces in many assembly applications such as toys. Sports applications such as fumble-free football gloves or rock climbing aids could be revolutionize sports equipment.

## Cross-References

► [van der Waals Forces](#)

## References

- K. Autumn, How gecko toes stick – the powerful, fantastic adhesive used by geckos is made of nanoscale hairs that engage tiny forces, inspiring envy among human imitators. *Am. Sci.* **94**(2), 124–132 (2006)
- K. Autumn, N. Gravish, Gecko adhesion: evolutionary nanotechnology. *Phil. Trans. R. Soc.- A Math. Phys. Eng. Sci.* **366**(1870), 1575–1590 (2008)
- H.J. Gao, H.M. Yao, Shape insensitive optimal adhesion of nanoscale fibrillar structures. *Proc. Natl. Acad. Sci. U. S. A.* **101**(21), 7851–7856 (2004)
- B.N.J. Persson, S. Gorb, The effect of surface roughness on the adhesion of elastic plates with application to biological systems. *J. Chem. Phys.* **119**(21), 11437–11444 (2003)
- M. Sitti, R.S. Fearing, Synthetic gecko foot-hair micro/nano-structures as dry adhesives. *J. Adhes. Sci. Tech.* **17**(8), 1055–1073 (2003)
- Y. Tian, N. Pesika, H.B. Zeng et al., Adhesion and friction in gecko toe attachment and detachment. *Proc. Natl. Acad. Sci. U. S. A.* **103**(51), 19320–19325 (2006)

## Generating Rack

► [Rack](#)

## Generic Solutions Due to a Unit Excitation

► [Influence Coefficients for Contact Mechanics](#)

## Geometry of Spherical/Aspheric Bearings

FENGCAI WANG<sup>1,2</sup>, HANYU WANG<sup>3</sup>

<sup>1</sup>School of Mechanical Engineering, Wuhan University of Science and Technology, Wuhan, Hubei, People's Republic of China

<sup>2</sup>National Research Centre of Bearing Technology (ZWZ), Xi'an Jiaotong University, People's Republic of China

<sup>3</sup>Bradford Grammar School, Bradford, West Yorkshire, UK

## Synonyms

EDSC – equivalent discrete spherical convolution considering spherical/aspheric geometry; ESC-equivalent spherical convolution considering spherical/aspheric geometry;



FTM – fixed-tracked method considering spherical/aspheric geometry; SFFT – spherical fast fourier transform considering spherical/aspheric geometry; SGDM – spherical grid data model considering spherical/aspheric geometry

## Definition

Geometry of spherical/aspheric bearing refers to the mathematical description of the shape and features of bearing surfaces for the purpose of understanding the tribological behavior of the bearings. Such geometrical description can be made in the different scales required for real tribological problems and the topographical characteristics of bearing surface, such as the macro- and micro-geometry of bearing surfaces. Despite the use of the term *spherical* in the formal definition, it must be noted that there are a number of factors that might contribute to the introduction of aspheric or non-spherical bearing geometry, such as tolerance and defect in manufacturing and wear in application. Additionally, compared with a perfectly spherical surface, both the specific surface design for a spherical bearing, such as surface texture, and other bearing forms, such as an ellipsoidal bearing, can lead to more complex topography. However, all aspheric bearings with either global or local non-sphericity can be studied through an equivalent spherical bearing modeling, which is particularly developed from spherical bearing modeling and thus falls sometimes into the same conceptual definition. The geometry definition is often linked to the development of geometry modeling for the study of tribological problems of spherical/aspheric bearing.

## Scientific Fundamentals

The descriptions of spherical/aspheric bearing geometry include various geometrical forms of spherical bearings, features of macro- or micro-geometry of bearings, representation of bearing geometry of spherical/aspheric or complex bearing in motion, and so on (Cheng et al. 2011, 2012a, b; Kim et al. 2006). Bearing geometry has a significant effect on tribological performances, such as deformation, contact mechanics, friction, wear, and lubrication (Wang et al. 2005a, b, c, 2006, 2008a, b, 2009a, b, c, 2010; Wang 2010, 2011). Therefore, the purpose of the recognition of various bearing geometries is related to their tribological problems under the given conditions, which are usually can be resolved by the corresponding spherical/equivalent spherical bearing simulation system. For this reason, the discrete geometry of spherical/aspheric bearings must be accurately described under the given set of operation conditions, such as a realistic three-

dimensional and time-dependent motion. A sufficient amount of geometry information is required for understanding how bearing geometry affects the tribological mechanism of bearings and is certainly useful for design, manufacturing, and application.

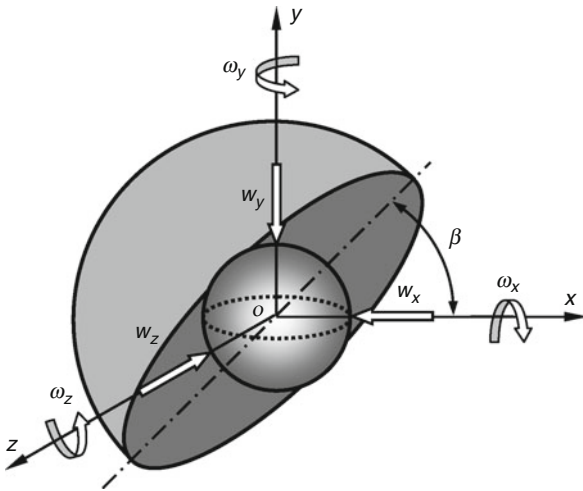
## Coordinate System and Spherical Mesh-Grids

The geometry of a spherical/aspheric bearing can be naturally described in the spherical coordinate system for the study of the corresponding tribological problems under a given loading and motion, as shown in Figs. 1 and 2 (Wang et al. 2008a). The chosen spherical coordinate system needs to be consistent with the spherical/aspheric bearing modeling required for solutions of the tribological problems. Since most tribological problems of spherical bearings are treated by numerical methods, the bearing geometry is discretely described in the spherical mesh-grid space. These are modeled by the development of the base-spherical bearing geometry model and hence the tribological solution of the spherical/aspheric bearing is usually named as the spherical-grid-data model (SGDM) (Wang et al. 2009a). In modeling a spherical/aspheric bearing, the geometry is meshed in the spherical coordinates  $(\varphi, \theta)$  with coordinate increments  $(\Delta\varphi, \Delta\theta)$ , respectively, as shown in Fig. 3. The position of a point  $p$  on the bearing surface is determined by its spherical coordinates  $(r, \varphi, \theta)$ .

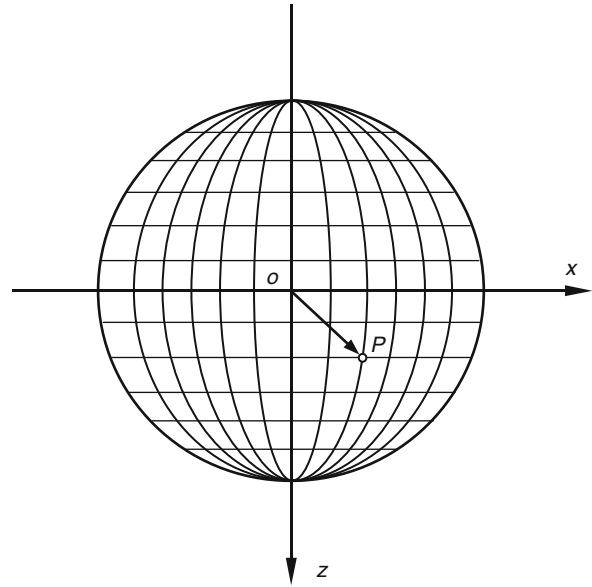
Sometimes the spherical/aspheric bearing may be positioned inclined rather than horizontally. For example, the spherical bearing used to represent both natural and artificial hip joints must be placed in a physiological position with an inclination angle of  $\beta$  as shown in Fig. 1 (Wang et al. 2005a). For tribological problems with an inclination angle, an appropriate spherical coordinate system should be chosen where the z-axis is always set through the two polar points of the spherical coordinates. The inclination angle of the spherical bearings can be readily considered by setting suitable boundary conditions for the solution domain, i.e.,  $\varphi = [\varphi_1, \varphi_2]$  and  $\theta = [\theta_1, \theta_2]$ , as shown in Fig. 4. Such a coordinate system for spherical bearing modeling including the bearing geometry inclination is referred as the Z-polar coordinate system (Wang et al. 2005b). This is also described in detail in the lubrication theory of spherical bearing and computational biotribology of hip joint replacements.

## Geometry of Spherical Bearing

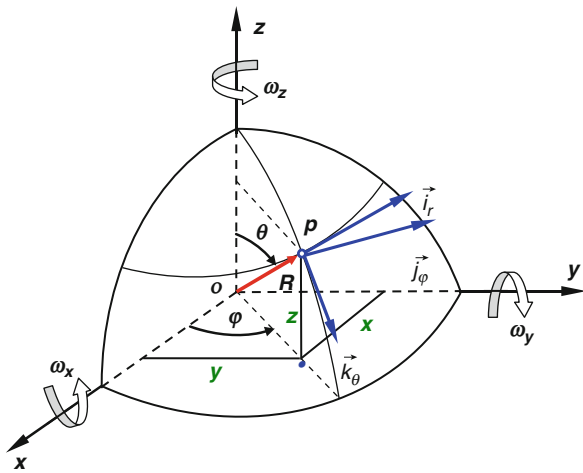
The geometry of bearing surfaces has a significant effect on the operational performance of the bearing. However, the description of bearing geometry is affected not



**Geometry of Spherical/Aspheric Bearings, Fig. 1** Schematic diagram of a spherical bearing under the three-dimensional loading and motion conditions

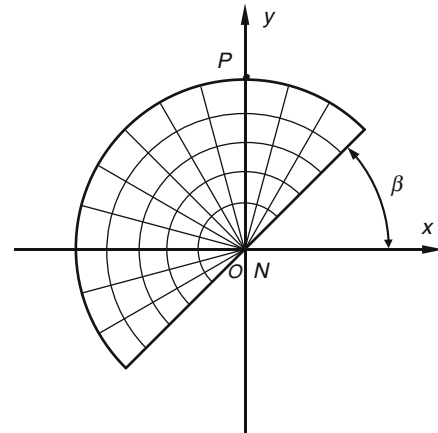


**Geometry of Spherical/Aspheric Bearings, Fig. 3** Spherical coordinate system and mesh grid for the geometrical description of spherical/aspheric bearings



**Geometry of Spherical/Aspheric Bearings, Fig. 2** Spherical coordinate system used for a spherical bearing model

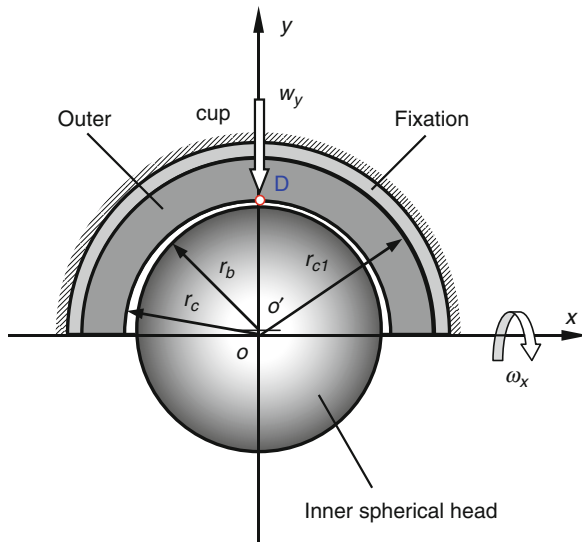
only by the geometrical features of the original bearing surfaces such as spherical or ellipsoidal bearing, but also by the geometrical changes due to bearing deformation, which are related to the operational conditions such as load, material property, and bearing structure. For example, the film thickness of elastohydrodynamic lubrication of bearings is given by the original bearing geometry and the elastic deformation of the bearing surfaces under the given operating conditions. Accurate



**Geometry of Spherical/Aspheric Bearings, Fig. 4**

An appropriate spherical coordinate system and mesh grids with the z-axis through the poles for a spherical bearing with an inclination angle  $\beta$

description of bearing geometry is thus one of the most important aspects of spherical/aspheric bearing modeling under both steady and time-dependent operating conditions (Wang et al. 2009b).



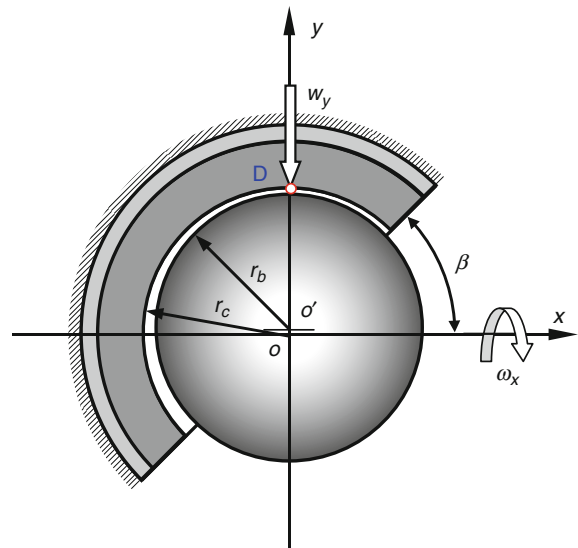
**Geometry of Spherical/Aspheric Bearings, Fig. 5**

A geometrical description of a spherical bearing positioned horizontally

A number of spherical/aspheric bearings with different forms of bearing structures are employed in general engineering and bioengineering applications. For example, a hip joint replacement comprises a spherical head bearing articulating against a hemi-spherical cup bearing, particularly as a conformal contact problem as shown in Fig. 5. The bearing geometry can be seen as either a perfect spherical or an aspheric surface. The accurate evaluation of the articulation between the head and the cup can be represented by typical ball-in-socket configurations and be applied to spherical bearing modeling for further tribological analysis (Wang et al. 2005a, 2009a). If the hemi-spherical cup is positioned with an inclination angle, the bearing geometry model can be described as shown in Fig. 6 where the spherical coordinates and the corresponding mesh-grid for the inclination problem can be chosen as shown in Fig. 4.

In fact, the spherical bearing has various forms. For example, the spherical bushing with the partial spherical surface as bearing surface is often used for general engineering applications for spherical bearings in both food machinery and hydraulic systems. The spherical bushings are self-aligning and have inner and outer rings with spherical bearing surfaces as shown in Fig. 7a, b.

The geometry of spherical/aspheric bearings can be described by using a perfect spherical surface as the base-spherical bearing surface. The geometry of spherical



**Geometry of Spherical/Aspheric Bearings, Fig. 6**

A geometry model of spherical bearing positioned with an inclination angle ( $\beta$ )

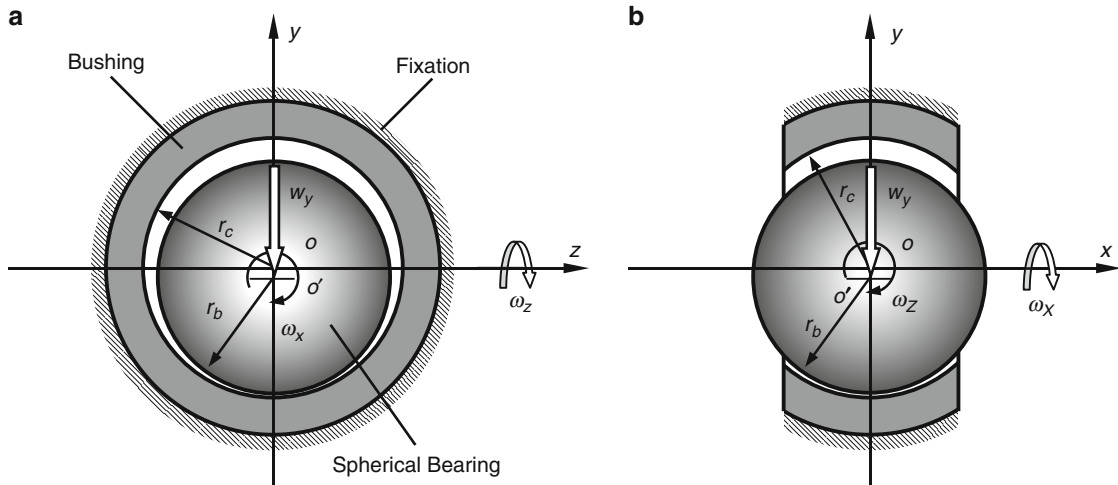
bearings can be represented by the radial distance from the original point in spherical coordinates as follows (Goenka et al. 1980; Wang et al. 2008a, b, 2009a, b, c):

$$r_i = r_{i0}(\varphi, \theta) \quad (1)$$

where the subscript  $i = c$  or  $b$  represents either the cup/bushing or the head/spherical bearing surfaces, which are usually in motion. Based on this bearing surface, both the global non-sphericity of aspheric bearings and the localized geometry changes can be modeled through the corresponding deviations from the nominal spherical bearing, such as the rough and the aspheric bearing surfaces.

### Global Non-sphericity of Bearing Geometry

There are a number of factors that may lead to more complex bearing surfaces with the non-sphericity of bearing geometry relative to the perfect spherical bearing, namely aspheric bearings (Wang et al. 2009a). As previously illustrated, nearly all the real spherical bearings are diversely non-spherical geometries due to manufacturing technology and operational application as well as possible specific design. The geometrical features of most aspheric bearings can be classified into two categories: one for the



**Geometry of Spherical/Aspheric Bearings, Fig. 7** A geometry model of a spherical bushing articulating against a spherical bearing: **(a)** viewed from the  $x$ -direction and **(b)** viewed from the  $z$ -direction

globalized non-sphericity, where the geometrical variations relative to the nominal spherical bearing geometry can be evaluated in the long/middle wavelength scale, such as the ellipsoidal bearings, and the other for the localized non-sphericity where the geometrical change features can be described in the short wavelength scale over the nominal spherical/smooth aspheric bearing geometry, such as roughness and surface texture. Both the global and the local non-sphericity of bearing geometry play an important role in the development of tribological modeling of spherical/aspheric bearings.

The overall geometry features due to the global non-sphericity of bearing geometry of real spherical/aspheric bearings can be found on the bearing surfaces, for example, the waviness and the ellipsoidal bearing geometry can be separately seen due to manufacturing and specific design. The aspheric bearings can be represented in the spherical coordinates by the deviations  $\Delta r_i(\varphi, \theta)$  of the aspheric bearing surface relative to a perfect spherical surface as the base-spherical surface in (1) as follows:

$$r_i = r_{i0}(\varphi, \theta) + \Delta r_i(\varphi, \theta) \quad (2)$$

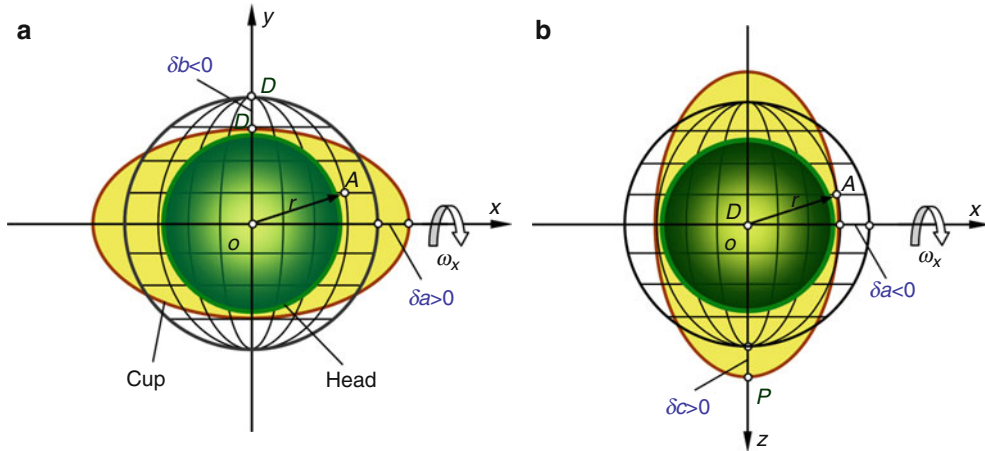
For example, taking an ellipsoidal bearing as an aspheric bearing with global non-sphericity of bearing geometry, it can be described by an ellipsoidal surface with variations of the three semi-axis lengths in Cartesian coordinates as follows (Wang et al. 2009a):

$$\frac{x^2}{(r_{i0} + \delta a)^2} + \frac{y^2}{(r_{i0} + \delta b)^2} + \frac{z^2}{(r_{i0} + \delta c)^2} = 1 \quad (3a)$$

where  $i = b$  or  $c$  for the inner and outer bearing surfaces, respectively, as shown in Fig. 8a, b. The variable  $r_{i0}$  represents the radius of the nominal spherical bearing surfaces. The variables  $(\delta a, \delta b, \delta c)$  are the deviations or the non-sphericity of the ellipsoidal bearing surface from the nominal spherical bearing surface in the  $x$ ,  $y$  and  $z$  directions, respectively. If all the non-sphericities  $\delta a$ ,  $\delta b$ , and  $\delta c$  are set to zero, the bearing surface becomes perfectly spherical (Wang et al. 2008a, b, 2009a, b).

In spherical coordinates, the radial distance at a point on the ellipsoidal bearing surface in (2) can be evaluated as follows:

$$r_i = \left[ \frac{(r_{i0} + \delta a)^2 (r_{i0} + \delta b)^2 (r_{i0} + \delta c)^2}{(r_{i0} + \delta b)^2 (r_{i0} + \delta c)^2 \cos^2 \varphi \sin^2 \theta + (r_{i0} + \delta a)^2 (r_{i0} + \delta c)^2 \sin^2 \varphi \sin^2 \theta + (r_{i0} + \delta a)^2 (r_{i0} + \delta b)^2 \cos^2 \theta} \right]^{\frac{1}{2}} \quad (3b)$$



**Geometry of Spherical/Aspheric Bearings, Fig. 8** Non-spherical geometry and mesh grid with the z-axis passing through the poles of the spherical coordinates: (a) with the variations of semi-axes  $\delta a$  and  $\delta b$  along x and y, and (b) with the variations of semi-axes  $\delta a$  and  $\delta c$  along x and z

where  $(r, \varphi, \theta)$  represent the spherical coordinates of the point on the ellipsoidal bearing surfaces.

### Local Non-sphericity of Bearing Geometry

Microscopic bearing geometry is certainly of non-spherical or freeform features. The micro-geometry of the bearing needs to be measured or described in the smaller scale with a high density mesh-grid, which is used for the corresponding tribological modeling of bearings (Wang et al. 2005b). The topographical features of micro-geometry of bearing surfaces such as rough or worn bearing surfaces are more complex in the geometry modeling than the global non-sphericity of bearing geometry. The real spherical/aspheric bearings under the consideration of the micro-geometrical variations can be expressed as follows:

$$r_i = r_{i0}(\varphi, \theta) + \Delta r_i(\varphi, \theta) + \Theta_i(\varphi, \theta) \quad (4)$$

where  $\Theta_i = \Theta(\varphi, \theta)$  is the micro-geometrical deviation of real bearing surface from the overall bearing geometry, such as roughness or micro machining marks. Compared with the global non-sphericity of bearing geometry, the micro-geometrical features might cause different effects on the tribological behavior of bearings.

### Deformation of Bearing Geometry

The change of bearing geometry can often be caused due to working under given loading and motion conditions since both the dynamic and the steady-state loading lead to the deformation of the bearing surfaces. Therefore, when the deformation is considered, the geometry of spherical/aspheric bearings can be further described by

$$r_i = r_{i0}(\varphi, \theta) + \Delta r_i(\varphi, \theta) + \Theta_i(\varphi, \theta) + \delta_i(\varphi, \theta) \quad (5)$$

Evaluating the deformation of spherical/aspheric bearings in spherical coordinates is not a trivial task. The changes of latitudes by spherical mesh-grid points result in the non-uniform mesh grids for the geometry description, particularly when a relative high density of mesh-grid points is required for the tribological modeling where a fast numerical technique has to be used. For example, the elastic deformation of spherical/aspheric bearings for the study of elastohydrodynamic lubrication is usually calculated by the equivalent spherical convolution model (ESC) or by the equivalent discrete spherical convolution model (EDSC) combined with the corresponding fast numerical solution techniques such as the spherical fast Fourier transform (SFFT) (Wang et al. 2004, 2008b, 2009a, b, c, d).

### Clearance and Curvature

For spherical/aspheric bearings, the radial clearance between the inner and outer spherical bearing surfaces is an important parameter in design and analysis that plays a significant role in the overall tribological performance (Wang et al. 2009b, d). It can be evaluated from the clearance of the corresponding nominal spherical bearing as follows:

$$C = r_{c0} - r_{b0} \quad (6)$$

Additionally, the change of curvature of the bearing surface is often used to study the tribological problems of the bearing. The curvatures at the point on bearing surface have maximum and minimum values, named as the principal curvatures,  $\kappa_1$  and  $\kappa_2$ , and their directions are called as the principal directions. Thus, Gaussian curvature and the mean curvature are defined as their geometric mean  $\kappa_g = \kappa_1 \kappa_2$  and arithmetic mean  $\kappa_m = (\kappa_1 + \kappa_2)/2$ , respectively. Being different from the perfectly spherical bearing, the curvatures could be changed at the different points due to the non-sphericity of the real spherical/aspheric bearings.

### Representation of Bearing Geometry

Real bearing surfaces are usually described by the discrete spherical mesh-grid space in spherical coordinates in order to carry out the simulation for the tribological problem of bearings, as shown in Fig. 1b. It is a key problem in the tribological modeling process that bearing surfaces are accurately described in the spherical mesh grid space for simulation by representing the bearing geometry in motion. On the other hand, the change of the bearing geometry due to wear with time in motion have to be remodeled over the corresponding mesh grids in the fixed coordinate system for simulation work. Such problems often cannot be directly resolved through the developed spherical grid data model (SGDM) since the bearing geometry is not perfectly spherical (Wang et al. 2009a, b, c, d). Therefore, the geometry representation of complex or moving bearing surfaces is required for the tribological modeling such as the lubrication problem of aspheric bearings and the bearing geometry modification of computational wear modeling. For example, the accurate geometrical representation needs to be made for the transient elastohydrodynamic lubrication problem in hip joint replacements with the non-sphericity of bearing geometry, where an extra transient effect might be caused by the non-spherical bearing geometry such as the ellipsoidal head bearing surface in motion (Wang et al. 2009a, b, d).

The key issue for the bearing geometry representation is how to effectively carry out a data communication between the bearing geometry in motion with time and the spherical mesh-grid space used for simulation. A fixed-tracked method (FTM) can be employed to finish the data communication of the remodeling of discrete bearing geometry (Wang et al. 2009a, b). The idea of the general methodology for data communication was built on the concept of building two sets of mesh-grids, one fixed in the space coordinates used for the numerical simulation and the other following the body coordinates

used for the description of the moving interface of bearing geometry in the corresponding tribological problem. The data communication between the two mesh-grid sets was carried out by an interpolation technique and the corresponding Euler angular transformation as follows:

$$[\mathbf{r}] = [\mathbf{A}]^{-1}[\mathbf{r}'] \quad (7a)$$

where the  $[\mathbf{r}]$  and  $[\mathbf{r}']$  are the location vectors at a point on the bearing surface in the space and body coordinate systems due to motion, which are given, respectively, in Cartesian and spherical coordinates by

$$[\mathbf{r}] = \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} r \sin \theta \cos \varphi \\ r \sin \theta \sin \varphi \\ r \cos \theta \end{Bmatrix} \quad (7b)$$

and

$$[\mathbf{r}'] = \begin{Bmatrix} x' \\ y' \\ z' \end{Bmatrix} = \begin{Bmatrix} r' \sin \theta' \cos \varphi' \\ r' \sin \theta' \sin \varphi' \\ r' \cos \theta' \end{Bmatrix} \quad (7c)$$

The inverse Euler angular transformation with three Euler angular motions in (7a) can be expressed as

$$A^{-1} = \begin{pmatrix} \cos \beta \cos \gamma & -\sin \beta & \cos \beta \sin \gamma \\ \cos \alpha \sin \beta \cos \gamma & \cos \alpha \cos \beta & \sin \alpha \cos \beta \\ \sin \alpha \sin \beta \cos \gamma & \sin \alpha \cos \beta & \sin \alpha \sin \beta \sin \gamma + \cos \alpha \cos \gamma \end{pmatrix} \quad (7d)$$

where  $(\alpha, \beta, \gamma)$  are three Euler angles corresponding, respectively, to the three Euler angular velocities  $(\omega_x, \omega_y, \omega_z)$  as shown in Fig. 1.

The general method can be applied to resolve not only the transient elastohydrodynamic lubrication of hip joint replacements where the moving interface can work under the conditions of the hard-on-hard material combination, the three-dimensional dynamic load, and the time-dependent velocity as well as the various non-sphericities of bearing geometry, but also other tribological problems under complex motion conditions, such as computational wear modeling and dynamic contact mechanics (Wang et al. 2009a, b, c, d).

### Film Thickness and Gap

Once the geometry description of spherical/aspheric bearing under the given conditions is accurately obtained through the above-mentioned geometrical modeling, the geometrical relationship between the two bearing surfaces can be determined by the lubricant film thickness



for the lubrication problem and the gap for the contact mechanics problem as follows (Wang et al. 2008a, b, 2009a, b, c, d):

$$h(\varphi, \theta, t) = (r_c - r_b) - e_x(t) \sin \theta \cos \varphi - e_y(t) \sin \theta \sin \varphi - e_z(t) \cos \theta + \Theta_{cb}(\varphi, \theta) + \delta_{cb}(\varphi, \theta, t) \quad (8)$$

where  $r_b$  and  $r_c$  are the radial distance from a point on the inner or outer bearing surfaces in the spherical/aspheric bearing to the center of the normal spherical bearing, respectively.  $(e_x, e_y, e_z)$  are the position of the center of the inner bearing surface with reference to the center of the outer bearing surface. They can also be described as a function of the eccentricity of the bearing,  $(\varepsilon_x, \varepsilon_y, \varepsilon_z)$ . The variables  $\Theta_{cb}$  and  $\delta_{cb}$  are the combined micro-geometrical features and deformation of the two bearing surfaces, respectively. When the steady-state problem is considered, the time-dependent term in the equation vanishes. The film thickness equation or the gap formula is usually employed for the study of both the lubrication modeling and the contact mechanics of spherical/aspheric bearing as a conformal contact problem.

## Key Applications

### Tribology of Spherical Bearing and Biotribology of Hip Joints

The description and modeling of spherical/aspheric geometry are widely applied to spherical/aspheric bearings in industrial engineering and bioengineering, such as the spherical bushing/joint in spatial flexible multi-body systems and hip joints. An accurate geometry description of bearing is significant due to the effect of the bearing geometry properties on the tribological performance of spherical bearings or biotribology of natural/artificial hip joints. A description of the overall geometry and the micro geometry as well as geometry changes due to deformations under the dynamic operational conditions can be used for the basic considerations of design and manufacturing as well as measurement in order to improve the tribological performance of bearings as much as possible. For example, the radial clearance of bearings is a basic design parameter, while the variation of the bearing geometry relative to the original spherical geometry due to either elastic deformation or non-sphericity plays a significant role in both steady-state and transient elastohydrodynamic lubrication. Therefore, a description of the macroscopic and the microscopic geometry can be employed to study the tribological problems of spherical/aspheric bearings in industrial engineering and biotribology of hip joints

in bioengineering (Wang et al. 2005a, b, c, 2008a, b, 2009a, b, c). A further explanation of application of geometry of spherical/aspheric bearings is made through the following aspects.

### Deformation and Contact Mechanics

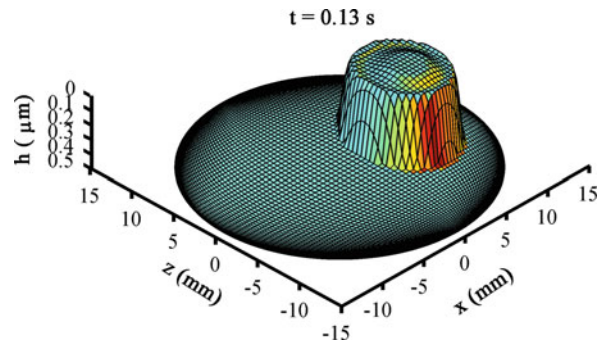
Both deformation and contact mechanics of spherical bearing surfaces are often used for the study of tribological problems such as lubrication, friction, and wear. All the simulation-based works need to accurately evaluate the changes of the bearing geometry and the gap between the bearing surfaces due to deformation and contact during loading and motion. A more accurate geometry modeling is also required for micro deformation and asperity contact mechanics, particularly when factors such as the micro-geometry features and the non-spherical bearing surface are considered (Wang et al. 2005c). The calculation of deformation of a spherical/aspheric bearing surface can be directly completed on the base-spherical surface as a spherical-grid-data-model by the EDSC model and SFFT technique (Wang et al. 2004, 2008a, b, 2009a, b, c, d). However, the simulation of deformation and the contact problems with either the spherical/aspheric bearing surfaces in motion or bearing geometrical modification due to wear can be evaluated by the fixed-tracked method (FTM). For example, the dynamic contact mechanics of a spherical bearing under the three-dimensional loading and motion conditions at any instant of time can be carried out through the representation of the bearing geometry in the mesh grid space ready for simulation by the method (Wang et al. 2009a). Therefore, the application of geometry modeling to deformation and contact is dependent on the requirements of the real problems (Wang et al. 2009a, b; Wang 2011). They play a significant role in the study of the tribological mechanism of spherical/aspheric bearings.

### Lubrication and Friction

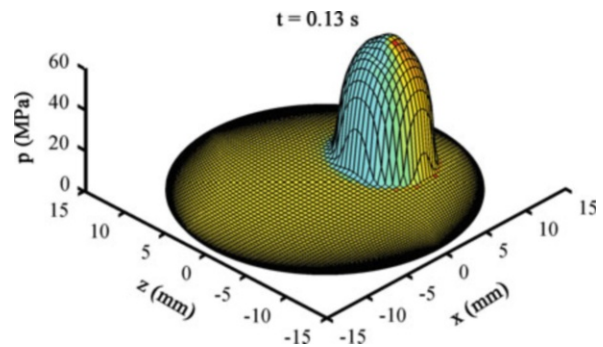
Lubrication and friction of spherical/aspheric bearings under given operating conditions can be simulated by using tribological modeling where the bearing geometry is required for an accurate description. Friction prediction is dependent on lubrication regimes such as hydrodynamic lubrication, elastohydrodynamic lubrication, and mixed/boundary lubrication (Wang et al. 2007, 2008b). Additionally, there are many topographical factors of both the global and local geometrical features that might have different effects on lubrication and friction of bearings, such as non-sphericity, roughness, and manufacturing defect (Wang et al. 2009b). The various geometrical features need to be effectively modeled to provide the

macroscopic and the microscopic geometry information for tribological modeling for the evaluation of the tribological performance of bearings. The geometry and the relevant description method, such as the above mentioned, can be employed to carry out the simulations of spherical/aspheric bearings. For example, an appropriate coordinate system with the  $z$ -polar axis must be chosen to simulate the lubrication and friction of hip joint replacement, which is physiologically positioned with an inclination angle on the cup bearing, while the main load is through the  $y$ -axis, as shown in Fig. 2. For the non-sphericity of bearing geometry, the representation of bearing geometry in motion can be carried out by the fixed-tracked method for the study of lubrication simulation of spherical/aspheric bearings. When the lubrication domain is a partial spherical surface such as a spherical bushing, the geometry used for the corresponding lubrication problem can be described by suitably defining the range of spherical coordinates ( $\varphi, \theta$ ). The lubrication and friction as well as contact problems often need to define either the film thickness or the gap between the two bearing surfaces. When the various geometrical components such as deformation, non-sphericity, and roughness are considered in the models, the whole geometry description can be carried out as given in (5). The effect of the elastic deformation on the spherical/aspheric bearing geometry can still be evaluated by the EDSC model for the elastic deformation and SFFT technique for the fast simulation in order to efficiently analyze the elastohydrodynamic lubrication (Wang et al. 2009a, b, c, d).

As an example of a spherical bearing, the transient elastohydrodynamic lubrication in a typical metal-on-metal hip joint replacement with the physiological inclination of the femoral cup bearing under walking conditions with dynamic loading and transient motion as well as a realistically low lubricant viscosity of 0.001 Pas is modeled, as shown in Figs. 9 and 10 (Wang et al. 2008a, b). The acetabular cup bearing is physiologically positioned with an inclination angle of  $45^\circ$ , as shown in Figs. 4 and 6. The predicted film thickness and pressure distributions at the time instant  $t = 0.13$  s of one walking cycle within a period of 1.0 s are evaluated. Additionally, the effect of the non-sphericity of the global bearing geometry of the hip joint replacement on the steady elastohydrodynamic lubrication can be evaluated through the general lubrication methodology of spherical/aspheric bearing with the fixed-tracked method for the representation of complex bearing geometry in motion. The fluid film thickness change in percent, the predicted pressure, and fluid film thickness distributions of the metal-on-metal hip joint replacement for the load



**Geometry of Spherical/Aspheric Bearings, Fig. 9** The film distribution at the instant 0.13 s of one walking cycle with a period of 1.0 s for a typical metal-on-metal hip joint replacement

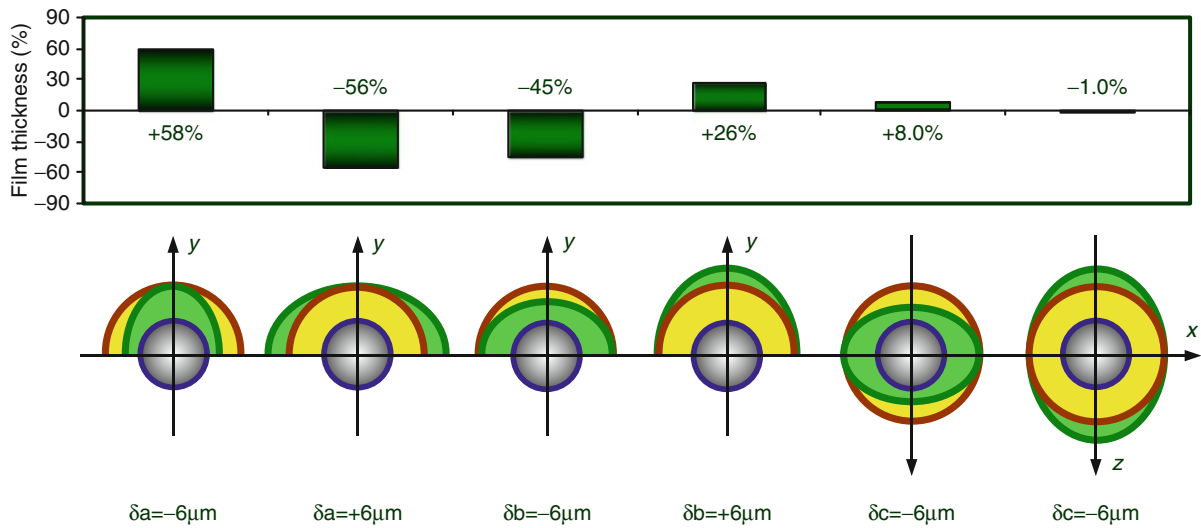


**Geometry of Spherical/Aspheric Bearings, Fig. 10** The predicted pressure at the instant  $t = 0.13$  s of one walking cycle with a period of 1.0 s for a typical metal-on-metal hip joint replacement

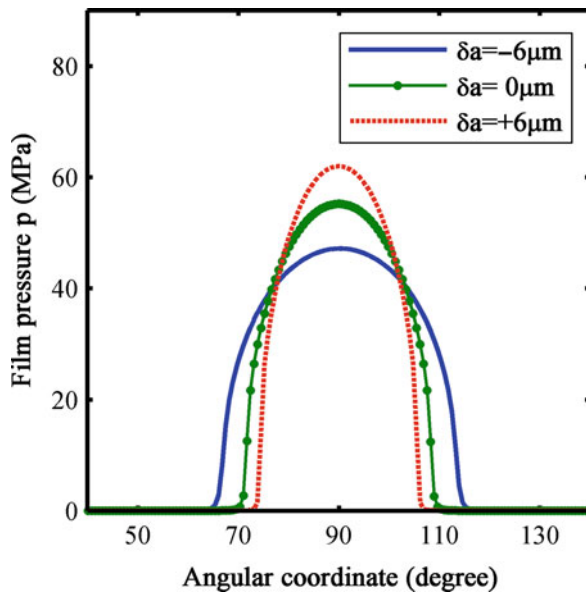
$w_y = 2,500$  N and the angular velocity  $\omega_x = -2$  rad/s under the different non-sphericities relative to the nominal spherical bearing along  $x$ -axis are shown in Figs. 11, 12, and 13, respectively (Wang et al. 2009a, b).

### Computational Wear Modeling

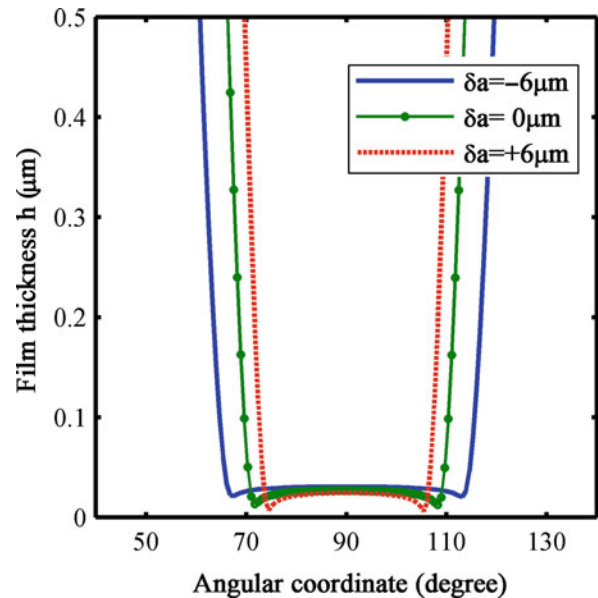
Wear processes always take place in bearings during operation. The geometry of the original bearing surfaces is changed with time due to wear. A computational model can be employed to simulate the wear process based on models such as Archard wear equation for either the total volumetric wear or the linear wear depth. The dynamic contact mechanics and wear factor need to be integrated in the wear model. However, the geometry modification of the bearing due to the wear process at any instant of time under dynamic loading and motion conditions has to be made in order to provide accurate geometrical



Geometry of Spherical/Aspheric Bearings, Fig. 11 The non-sphericity of ellipsoidal cup bearing along the three semi-axis directions and the change of the corresponding film thickness



Geometry of Spherical/Aspheric Bearings, Fig. 12 Effect of non-sphericity of bearing geometry on elastohydrodynamic lubrication: film pressure profiles through the center of cup bearing with the different non-sphericity  $\delta a$  along the x-axis in  $\phi$  direction



Geometry of Spherical/Aspheric Bearings, Fig. 13 Effect of non-sphericity of bearing geometry on elastohydrodynamic lubrication: film thickness profiles through the center of cup bearing with the different non-sphericity  $\delta a$  along the x-axis in  $\phi$  direction

information for the wear modeling. The geometry modification can be finished by using the fixed-tracked method (FTM) for the geometrical representation of the worn bearing surfaces at any instant of time in the spherical

mesh-grid space for the wear simulation (Hu et al. 2011; Harun et al. 2007; Wang et al. 2009b, c; Wang 2010). For instance, the geometrical change of the two bearings surfaces due to wear with time takes place in hip joint

replacements with the hard-on-hard material combination; the geometrical representation of femoral head bearing in motion into the spherical mesh-grid space for simulation can be made by the fixed-tracked method. Thus, the femoral head bearing surfaces modified due to the wear process can be represented at any instant of time on the fixed spherical mesh grid space, while the other mesh grid space can follow the femoral head bearing surface in motion and transfer their geometrical information into the fixed mesh grid space. All simulations for the computational wear modeling are carried out in the fixed spherical mesh grid space.

For the specific details of the bearing geometry used for the study of both the tribological mechanism of spherical/aspheric bearings and biotribology of hip joints, refer to the following entries: ► [Elasticity Theory for Spherical Bearings](#), ► [Contact Mechanics for Spherical/Aspheric Bearing](#), ► [Lubrication Theory for Spherical Bearings](#), ► [Friction Prediction for Spherical Bearings](#), and ► [Wear Modeling of Spherical Bearings](#)

## Cross-References

- [Contact Mechanics for Spherical/Aspheric Bearing](#)
- [Elasticity Theory for Spherical Bearings](#)
- [Friction Prediction for Spherical Bearings](#)
- [Lubrication Theory for Spherical Bearings](#)
- [Wear Modeling of Spherical Bearings](#)

## References

- Q. Cheng et al., Effect of edge contact behavior of metal-on-metal hip joint replacements. *J. Med. Biomechan.* **26**(6), 540–545 (2011)
- Q. Cheng et al., Dynamic edge contact behavior of metal-on-metal hip joint replacements. *J. Med. Biomechan.* **27**(2), 29–34 (2012a)
- Q. Cheng et al., Contact mechanics of non-spherical hip joint replacements. *J. Med. Biomechan.* **27**(5), 534–544 (2012b)
- P.K. Goenka et al., Spherical bearings: static and dynamics analysis via the finite element method. *ASME J. Lubr. Technol.* **102**(7), 308–319 (1980)
- M.N. Harun et al., Development of computational wear simulation of metal-on-metal hip joint replacements. *Transactions of the Annual Meeting of the Orthopaedic Research Society*, 32, Paper No. 1661, 2007
- Z.M. Hu et al., Wear prediction of metal-on-metal hip joint replacements. *J. Med. Biomechan.* **26**(5), 448–453 (2011)
- B.C. Kim et al., Development of composite spherical bearing. *Compos. Struct.* **75**, 231–240 (2006)
- F.C. Wang et al., Prediction of elastic deformation of acetabular cup and femoral head for lubrication analysis of artificial hip joints. *Proc. Inst. Mech. Eng. J. J. Eng. Tribol.* **218**, 201–208 (2004)
- F.C. Wang et al., Elastohydrodynamic lubrication modeling of artificial hip joints under steady-state conditions. *ASME J. Tribol.* **127**(10), 729–739 (2005a)
- F.C. Wang et al., Elastohydrodynamic lubrication modeling of spherical metal-on-metal artificial hip joints, in *ASME Proceedings of WTC2005, World Tribology Congress III*, Washington, DC, 2005b, WTC2005-63556, pp. 489–490
- F.C. Wang et al., An integrated experimental and theoretical contact mechanics study of UHMWPE hip implants tested in a hip simulator, in *ASME Proceedings of World Tribology Congress III*, Washington, DC, 2005c, pp. 311–312
- F.C. Wang et al., Lubrication modeling of artificial hip joints, in *IUTAM Symposium on Elastohydrodynamic and Micro-Elastohydrodynamic 134: Book Series: Solid Mechanics and Its Applications*, ed. by R.W. Snidle, H.P. Evans, Springer, 2006, pp. 385–396
- F.C. Wang et al., Effect of non-spherical bearing geometry on transient elastohydrodynamic lubrication in metal-on-metal hip joint implants. *Proc. Inst. Mech. Eng. J. J. Eng. Tribol.* **221**, 379–389 (2007)
- F.C. Wang et al., Transient elastohydrodynamic lubrication of hip joint implants. *ASME J. Tribol.* **130**(1), p011007 (2008a)
- F.C. Wang et al., Lubrication and friction prediction in metal-on-metal hip joint implants. *IOP Phys. Med. Biol.* **53**, 1277–1293 (2008b)
- F.C. Wang et al., Nonsphericity of bearing geometry and lubrication in hip joint implants. *ASME J. Tribol.* **131**(3), p031201 (2009a)
- F.C. Wang et al., Non-spherical bearing geometry and elastohydrodynamic lubrication of hip joint replacements under transient walking conditions, in *World Tribology Congress IV*, Kyoto, 2009b
- F.C. Wang et al., Dynamic contact mechanics and wear modeling of hip joint replacements with hard-on-hard material combination under three-dimensional loading and transient motion, in *World Tribology Congress IV*, Kyoto, 2009c
- F.C. Wang et al., Non-sphericity of bearing geometry and lubrication in hip joint implants. *ASME J. Tribol.* **131**(3), p031201 (2009d)
- F.C. Wang et al., Tribological modeling of spherical bearing with complex spherical-base geometry and motion. *Tribology and Design 2010*, Algarve (2010)
- F.C. Wang, Dynamic contact behaviour and evolution of bearing interface with spherical-base geometry, National Natural Science Foundation of China (NSFC) Report No. 10972165, pp. 1–30 (2010)
- F.C. Wang, Dynamic contact mechanism and failure of rolling bearings, national key basic research program of China (973 Program) Report No. 2011CB706601 (2011)

## Giga Pascal (GPa)

- [Plasto-Elastohydrodynamic Lubrication \(PEHL\)](#)

## Gigacycle Fatigue (GCF)

- [Super-Long Life Fatigue](#)

## Glass Transition

- [Temperature and Pressure Dependence of Density and Thermal Conductivity of Liquids](#)

## Glass-Forming Ability (GFA)

- Crack Growth in Noncrystalline Solids

## Glass-Transition Temperature ( $T_g$ )

- Crack Growth in Noncrystalline Solids

## Glaze Layer

- Self-Mating Metal Articulations in the Hip Joint

## Glide Testing of Magnetic Disk

- Disk Roughness and Defect Monitoring

## Glowing Contacts

XIN ZHOU

Innovation Center, Eaton Corporation, Moon Township, PA, USA

### Synonyms

Overheated contacts

### Definition

Glowing contacts are overheated electric contacts that have a red glow due to high temperature induced by Joule heating at the contact interface.

### Scientific Fundamentals

Glowing contacts are electric contacts that have high contact resistance at the contact interfaces. The high contact resistance leads to strong Joule heating and results in overheated electric contacts (Meese and Beausoliel 1977; Sletbak et al. 1992; Shea 2006; Shea and Zhou 2009; Zhou et al. 2008). For an electric contact interface that has a contact spot with equivalent radius much smaller than the contact surface size, a simplified expression can be used to describe the relationship between the contact

constriction resistance and the contact force, electrical resistivity of contact material, and the hardness of contact material (Sletbak et al. 1992).

$$R_c = \frac{\rho}{D} = \frac{\rho}{2} \sqrt{\frac{\pi H}{F}}$$

where  $R_c$  is the contact constriction resistance [ $\Omega$ ],  $\rho$  is the contact material resistivity [ $\Omega/\text{m}$ ],  $F$  is the contact force [ $\text{N}$ ], and  $H$  is the contact material hardness [ $\text{N}/\text{m}^2$ ].

Joule heating due to the constriction resistance across the contact interface can be expressed by the following equation:

$$W = R_c I^2$$

where  $I$  is the current [ $\text{A}$ ].

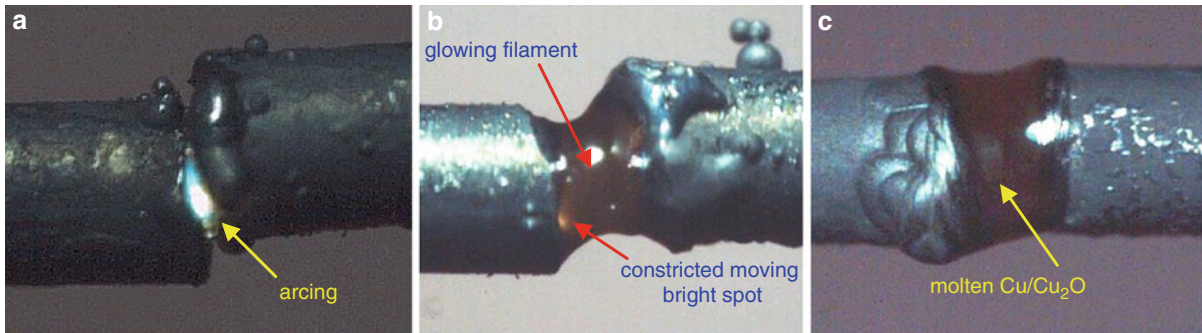
According to Wiedemann-Franz law, the following equation describes the relationship between the voltage drop across the contact interface and contact surface temperature increase ( $T_I - T_b$ ):

$$T_I = \sqrt{T_b^2 + \frac{\Delta V_c^2}{4L}}$$

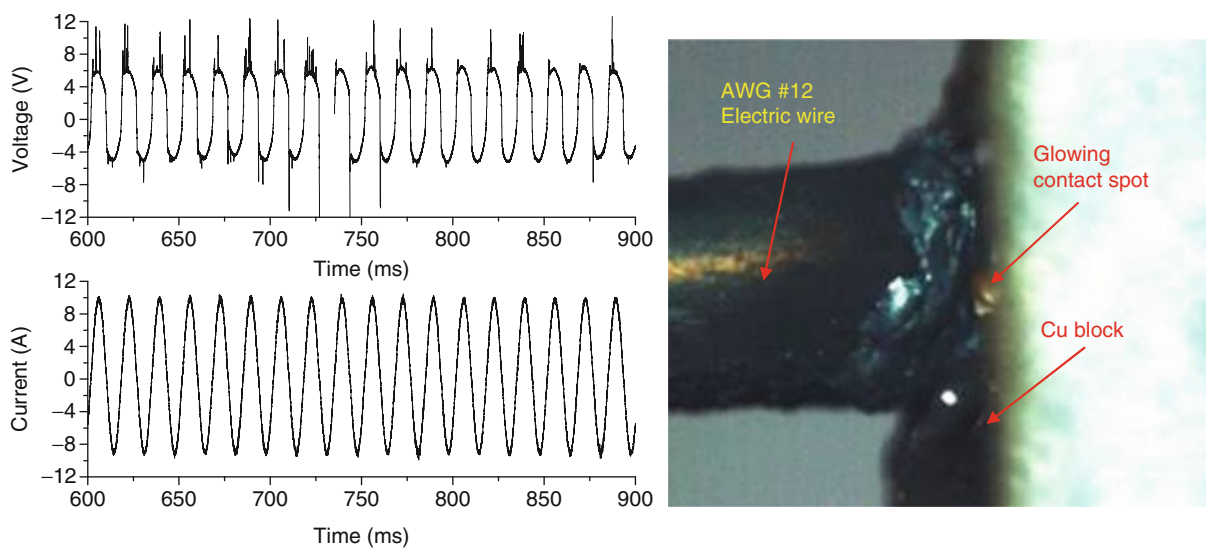
where  $T_I$  is the maximum temperature at contact interface [ $\text{K}$ ],  $T_b$  is the bulk temperature [ $\text{K}$ ] of the contact, and  $L$  is Lorentz constant ( $2.45 \times 10^{-8} \text{ V}^2/\text{K}^2$ ). The temperature of a glowing contact ranges from hundreds of degrees to thousands of degrees at the melting or boiling temperature of the metal or metal oxide (Shea 2006). Usually, the interface at the electric joints turns glowing red due to the high temperature.

Different modes at contact interfaces can exist during the formation process of glowing contacts. Usually, it takes a long time to form glowing contacts. The process typically starts with oxidation and/or corrosion at the contact interface. As the percentage of metal oxides at the contact interface increases, the contact constriction resistance increases. Joule heating due to the high contact resistance in turn accelerates the oxidation and/or the corrosion process. This further increases contact resistance and Joule heating at the contact interface, which leads to the formation of glowing contacts. During the formation process of glowing contacts, there can be intermittent arcing due to the rupture of the molten contact spot at the interface. A molten, red, glowing constricted current conducting path can also form and move around at the contact interface. The glowing contact resistance increases as the current decreases (Shea 2006; Shea and Zhou 2009). For example, in the case of two opposing thin wire electrodes, the voltage drop across the glowing contacts can be as high as 9.6 V for phosphor bronze contacts when the current is 0.5  $A_{\text{rms}}$ .





**Glowing Contacts, Fig. 1** Glowing contacts in different modes (Source Zhou et al. 2008)



**Glowing Contacts, Fig. 2** Voltage and current traces of a glowing contact in Mode I formed at the interface between a 1 mm diameter Cu wire and a Cu connector (Source Zhou et al. 2008)

Figure 1 shows three different modes during the formation process of glowing contacts (Zhou et al. 2008). In this case, the glowing contacts were generated by using two opposing 1 mm diameter Cu wires. Mode I (a) typically occurs at currents of 5 Amps or higher. Due to Joule heating and instability of the molten  $\text{Cu}_2\text{O}$ , the glowing contact would momentarily rupture into an arc, and then re-forms again. Mode II (b) is characterized by moving glowing filaments. This typically occurs at currents below 5 Amps. In this mode, a constricted moving bright spot and glowing filament bridging the two opposing Cu-wires can be observed. Mode III (c) is the completely molten state. There is no constricted moving bright spot, no moving filament and no arcing. This mode typically occurs after Mode I or Mode II has formed with sufficient  $\text{Cu}_2\text{O}$  at the contact interface.

Figure 2 shows a glowing contact in Mode I formed at the interface between a 1 mm diameter Cu wire and a Cu connector. The voltage trace indicates that the intermittent arcing occurred almost at the peak of the current.

### Key Applications

Glowing contacts are one of the major precursors for electrical fire and arc fault. Figure 3 shows a glowing contact connection in an electrical switch. Due to its relatively low impedance compared with system voltages in residential, commercial, and industrial applications, it is very difficult to detect an overheated or glowing contact (usually associated with loose electrical connections) by monitoring voltage and current signals of the electrical system. Common technologies





**Glowing Contacts, Fig. 3** A glowing contact connection in an electrical switch (Courtesy of Eaton Corporation)

employed today for detecting overheated contacts or glowing contacts are infrared thermal imaging and temperature sensing.

### Cross-References

- [Contact Boiling Voltage](#)
- [Contact Melting Voltage](#)
- [Contact Temperature of a Moving Solid Surface](#)
- [Contact Temperature of a Stationary Solid Surface](#)
- [Contact Temperatures on Coated or Rough Solid Surfaces](#)

### References

- W.J. Meese, R.W. Beausoliel, *Exploratory study of glowing electrical connections*, NBS building science series 103, Oct 1977
- J. Shea, Glowing contact physics, in *Proceedings of IEEE Holm conference*, 2006
- J. Shea, X. Zhou, Material effect on glowing contact properties. *IEEE Trans. CPT* **32**(4), 734–740 (2009)
- J. Sletbak, R. Kristensen, H. Sundklakk, G. Navik, M. Runde, Glowing contact areas in loose copper wire connections. *IEEE Trans. CHMT* **15**(3), 322–327 (1992)
- X. Zhou, J. Hastings, T. Schoepf, Detection of glowing contacts using acoustic sensing technology, in *Proceedings of the 24th international conference on electric contacts*, 2008

## Governing Equations of Fluid Motion

- [Navier-Stokes Equation and Applications in Lubrication](#)

## Gram Force (gf)

- [Sputtering MoS<sub>2</sub>-based Coatings](#)

## Granular Lubrication

M. M. KHONSARI

Department of Mechanical Engineering, Louisiana State University, Baton Rouge, LA, USA

### Definition

Alternative lubricants made of fine granules for applications where conventional lubricants fail.

### Scientific Fundamentals

The fundamental beauty and wide range of applicability of granular flow for predicting the behavior of diverse physical phenomena – both natural and engineered systems – has captured the attention of many prominent scientists and engineers for well over a century. Interestingly, before publishing his celebrated work on hydrodynamic lubrication theory, Reynolds (1885) wrote one of the first papers that described the dilatation tendency of compacted granular material when subjected to shearing action. More than a half-century later, Bagnold (1954) published the results of his pioneering experimental studies of the flow of granular material under shear that revealed distinct flow patterns — the so-called granule-inertia and macroviscous — associated with dispersed solid spherical granules of uniform size that were sheared between two concentric drums. Research on the flow characteristics of granular materials has continued steadily since then. Within the context of tribology, however, the subject has received increasing attention only over the past two decades (Elrod 1988; Elrod and Brewe 1991).

### Key Application

A major motivation has been the need to lubricate machinery operating under severe conditions. Integrated High Performance Turbine Engine Technology (IHPTET) and Versatile Affordable Advanced Turbine Engines program (VAATE) call for very high operating temperatures, in the neighborhood of 800°C or more (Dareing and Dayton 1992; Valenti 1995). Given that most conventional oils tend to break down when exposed to temperatures over 200°C, these applications pose very serious challenges for lubrication of vital components such as bearings.

In searching for an alternative lubricant, Heshmat (1992) demonstrated that effective lubrication can be achieved by injecting granular powder made of  $\text{TiO}_2$  into the clearance space of a thrust bearing. Remarkably, he obtained pressure profiles that were similar to that of a liquid lubricated thrust bearing. Later investigations (Heshmat 1994) led to successful development of the powder-lubricated bearings (Heshmat and Brewe 1995) that can operate at  $2 \times 10^6$  DN (bearing diameter in mm  $\times$  bearing speed in rpm). What followed these developments was a host of research programs in the tribology community — analytical and experimental — devoted to better understanding of the nature of granular lubrication (See for example, Dai et al. 1994; Yu et al. 1994; Yu and Tichy 1996; Zhou and Khonsari 2000; Sawyer and Tichy 2004; Kaur 2001; Kuar and Heshmat 2002; Higgs and Tichy 2004; Iordanoff and Khonsari 2004).

## Modeling Aspects

A comprehensive review is given by Worniyoh et al. (2007) with appropriate discussion on the similarities and differences between *powder* lubrication — characterized by cohesive soft particles on the order of microns — and *granular* lubrication — characterized by cohesionless hard particles with sizes on the order of millimeters. There are, indeed, subtle differences insofar as the lubrication mechanism and modeling aspects are concerned. Yet a thorough discussion is not within the scope of this essay, and the interested reader is referred to Worniyoh et al. (2007) and references contained therein.

It is, nevertheless, important to note the differences between the features of liquid and granular lubrication. Some of the major differences are:

- Conventionally, liquid lubrications conform to the no-slip condition at the boundaries. If the surface is stationary, the liquid at that boundary remains motionless, and if the surface is sliding, so is the liquid attached to it. The no-slip condition does not necessarily hold when dealing with granular materials; granules slip at the boundaries. The amount of “slippage” is not a priori known. However, there are analytical methods for the estimation of the magnitude (Huff and Hui 1984; Jenkins and Richman 1986; McKague and Khonsari 1996a).
- Collision of granules and their rebounding against solid boundaries is important in characterizing granular lubrication behavior, whereas this is absent in liquid lubrication.
- In granular flows, surfaces as a whole play a more active role on the performance of granular materials than when operating with liquid lubricants. Surface features (roughness) are considered essential in granular lubrication modeling, albeit with a simple coefficient of specularity that varies between 0 (ultra smooth) to 1 (ultra rough). The scale of roughness is typically on the order of the size of the granules.
- Theoretically, as predicted by the Reynolds hydrodynamic equation, two parallel disks in relative sliding motion and containing linearly viscous (i.e., Newtonian) liquid cannot support external loads. In contrast, measurable lift can be predicted with granular materials between parallel disks and interpreted as load-carrying capacity (cf. El Kholy and Khonsari 2008a).
- An important aspect of granular materials flow with relevance to tribology is a peculiar characteristic known as the *jammed state* in granular media, where the internal structure resists stress, forces do not propagate uniformly, and granules localize along directional force chains. The successive formation and collapse of jammed states causes fluctuations and results in formation of stick-slip (Albert et al. 2001). These effects figure importantly in the study of avalanches (Johansen et al. 1993; Pöschel and Buchholtz 1993) and earthquakes (Morales et al. 1996). Within the context of tribology, the spatial arrangement of granules and associated stick-slip affect the friction behavior of granular lubrication, particularly at low speeds (El Kholy and Khonsari 2008b).

## Theoretical Background

Several reviews of analytical modeling aspects are available (see, for example, Elrod 1988; Khonsari 1997; Worniyoh et al. 2007). One approach is based on the general theory of granular theory developed by Haff (1983). Motivated by molecular motion within a dense gas, the structure of Haff’s formulation was based on the corresponding kinetic theories. The application of the Haff’s theory to lubrication problems has been the subject of a number of publications (see, for example, Dai et al. (1994); McKague and Khonsari 1996a, b; Tsai and Jeng 2006). These analyses show that the physics of granular lubrication can be properly captured by the phenomenological relationships developed by Haff. In that theory, particles do not possess an enduring contact, which implies that the stress field is governed entirely by the collision and kinetic stress. A more comprehensive theory of granular material is developed by Johnson and Jackson (1987) that uses the Lun and colleagues’ (1984) constitutive equation and accounts for effect of sliding

motion and colliding effects of particles. A brief description of the appropriate equations follows.

### Conservation Equations

The conservation laws involve the conservation of mass, momentum, and pseudo-energy representing particle fluctuations. These are reviewed below. If the actual temperature is desired, then the “true” temperature would have to be considered as well.

*Conservation of mass.* The governing equation for the conservation of mass is

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{U}) = 0 \quad (1)$$

where  $\rho = \rho_p v$  is the bulk density of the granular material,  $\rho_p$  is the density of the individual granule,  $v$  is the solid volume fraction, and  $\vec{U}$  is the bulk velocity.

*Conservation of momentum.* The conservation of momentum is governed by

$$\rho \frac{D\vec{U}}{Dt} = \rho \vec{g} - \nabla \cdot \tilde{\sigma} \quad (2)$$

where  $\vec{g}$  is the gravity acceleration. The total stress is  $\tilde{\sigma} = \tilde{\sigma}_{ck} + \tilde{\sigma}_f$ , where  $\tilde{\sigma}_{ck}$  is the stress tensor caused by the collision and the kinetic motion and  $\tilde{\sigma}_f$  is the frictional stress tensor due to the enduring contact.

*Conservation of pseudo-energy energy.*

$$\frac{3}{2} \rho \frac{DT}{Dt} = -\nabla \cdot \vec{q}_{pT} - \tilde{\sigma}_{ck} : \nabla \vec{U} - \gamma \quad (3)$$

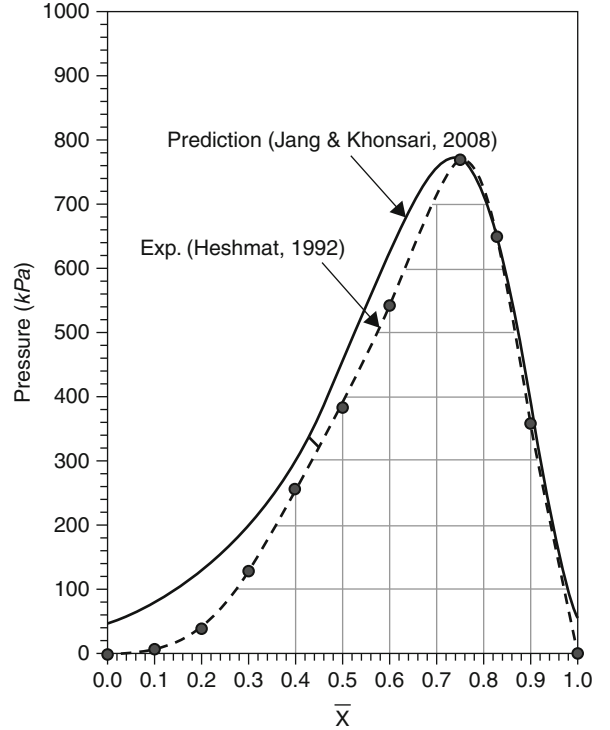
where  $T = v_m^2/3$  is the so-called *pseudo-thermal temperature* defined in terms of the mean fluctuation velocity  $v_m$  of the granule.  $\vec{q}_{pT}$  is the flux of the pseudo-thermal energy,  $\tilde{\sigma}_{ck} : \nabla \vec{U}$  is the rate of work by the collisional-translational component of the stress, and  $\gamma$  is the rate of energy dissipation due to the inelastic collision between granules

### Constitutive Equations

The stress tensor due to collision and kinetic motion is described by (Lun et al. 1984):

$$\begin{aligned} \tilde{\sigma}_{ck} = & [\rho T (1 + 4\eta v g_0) - \eta \mu_b \nabla \cdot \vec{U}] \tilde{I} \\ & - \left[ \frac{2\mu}{\eta(2-\eta)g_0} \left\{ 1 + \frac{8}{5} v \eta g_0 \right\} \left\{ 1 + \frac{8}{5} \eta (3\eta - 2) v g_0 + \frac{6}{5} \eta \mu_b \right\} \tilde{S} \right] \end{aligned} \quad (4)$$

where  $\tilde{I}$  is the identity tensor and  $\tilde{S}$  is the deviatoric part of the rate of deformation tensor. The parameter  $g_0$  represents the radial distribution function proposed by



**Granular Lubrication, Fig. 1** Pressure distribution in a slider bearing lubricated with  $\text{TiO}_2$

Canahan and Starling (1969). Parameter  $\eta$  is a constant characterizing the inelastic collision between granules. Parameters  $\mu_b$  and  $\mu$  are the bulk viscosity and the shear viscosity for perfectly elastic particles, respectively.

### Typical Pressure Distribution in a Bearing

These conservation laws were used by Zhou and Khonsari (2000) and later by Jang and Khonsari (2005) to develop a three-dimensional generalized Reynolds-type equation that predicts the pressure profile in a bearing with any specified film profile. This is particularly useful for prediction of the performance of granular lubricated bearings. Figure 1 shows the results of prediction of the Reynolds equation along with comparison to the experimental results of Heshmat (1992). Close agreement shows the applicability of the approach to granular lubrication. Later, the theory was applied to study the enduring contact between granules of powder lubricant in a bearing (Jang and Khonsari 2006). It is found that the enduring contact tends to dominate the kinetic regime of a powder lubricant at low speed. Research activities continue to explore the implications of these findings in various tribological applications.

## References

- I. Albert, P. Tegzes, R. Albert, J.G. Sample, A.-L. Barabási, T. Vicsek, B. Kahng, P. Schiffer, Stick-slip fluctuations in granular drag. *Phys. Rev. E* **64**(031307), 1–9 (2001)
- R. Bagnold, Experiments on a gravity-free dispersion of large solid spheres in a Newtonian fluid under shear. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* **225**(1160), 45–63 (1954)
- F. Dai, M.M. Khonsari, Y.Z. Lu, On the lubrication mechanism of grain flows. *Soc. Tribol. Lubr. Eng. (STLE) Tribol. Trans.* **37**, 516–524 (1994)
- D. Dareing, R. Dayton, Non-Newtonian behavior of lubricant mixed with ethylene glycol. *STLE Tribology Transactions*, **35**, 114–120 (1992)
- K. Elkholy, M.M. Khonsari, Granular collision lubrication: Experimental investigation and comparison with theory. *ASME J. Tribol.* **129**, 923–932 (2007)
- K. Elkholy, M.M. Khonsari, On the effect of enduring contact on the flow and thermal characteristics in powder lubrication. *Proc. IMechE Part J J. Eng. Tribol.* **222**, 741–759 (2008a)
- K. Elkholy, M.M. Khonsari, Experimental investigation on the stick-slip phenomenon in granular collision lubrication. *ASME J. Tribol.* **130**, 021302 1–7 (2008b)
- H. G. Elrod, Granular flow as a tribological mechanism – A first look, in *Interface Dynamics, Proceedings of the Leeds-Lyon Conference* Butterworth, Oxford, pp. 75–102 (1988)
- H.G. Elrod, D.E. Brew, Numerical experiments with flows of Elongated granules, in *18th Leeds Lyon Symposium on Tribology*. Elsevier Tribology Series, vol. 21, pp. 219–226 (1991)
- P.K. Haff, Grain flow as a fluid-mechanical phenomenon. *J. Fluid Mech.* **134**, 401–430 (1983)
- H. Heshmat, High-temperature solid-lubricated bearing development – Dry powder-lubricated traction testing. *J. Propul. Power* **75**, 814–820 (1991)
- H. Heshmat, The quasi-hydrodynamic mechanism of powder lubrication – Part II. Lubricant film pressure profile. *J. STLE Lubr. Eng.* **48**(5), 373–383 (1992)
- H. Heshmat, D. Brew, Performance of powder-lubricated journal bearings with MoS<sub>2</sub> powder: Experimental study of thermal phenomena. *ASME J. Tribol.* **117**, 506–512 (1995)
- H. Heshmat, On the theory of quai-hydrodynamic lubrication with dry powder: Application to development of high-speed journal bearings for hostile environments, in *20th Leeds-Lyon Symposium on Dissipative Processes in Tribology*. Tribology series, vol 23 (Elsevier Science, Lyon, 1994)
- C.F. Higgs III, J. Tichy, Granular flow lubrication: Continuum modeling of shear behavior. *ASME J. Tribol.* **126**, 499–510 (2004)
- K. Hui, P.K. Haff, J.E. Ungar, R. Jackson, Boundary conditions for high-shear grain flows. *J. Fluid Mech.* **145**, 223–233 (1984)
- I. Iordanoff, M.M. Khonsari, Granular lubrication: Towards an understanding of the transition between kinetic and quasi-fluid regime. *ASME J. Tribol.* **126**, 137–145 (2004)
- I. Iordanoff, B. Seve, Y. Berthier, Solid third body analysis using a discrete approach: Influence of adhesion and particle size on the macroscopic behavior of the contact. *ASME J. Tribol.* **124**, 530–538 (2002)
- J.Y. Jang, M.M. Khonsari, On the granular lubrication theory. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* **461**(2062), 3255–3278 (2005)
- J.Y. Jang, M.M. Khonsari, On the role of enduring contact in powder lubrication. *ASME J. Tribol.* **128**(1), 168–175 (2006)
- J.T. Jenkins, M.W. Richman, Boundary conditions for plane flows of smooth, nearly elastic, circular disks. *J. Fluid Mech.* **171**, 53–69 (1986)
- A. Johansen, P. Dimon, C. Ellegaard, J.S. Larsen, H.H. Rugh, Dynamic phases in a spring-block system. *Phys. Rev. E* **48**(6), 4779–4790 (1993)
- P.C. Johnson, R. Jackson, Frictional-collisional constitutive relations for granular materials with applications to plane shearing. *J. Fluid Mech.* **176**, 67–93 (1987)
- R. Kaur, The power of powder. *Lubs-n-Greases* **7**(5), 50–54 (2001)
- R.G. Kaur, H. Heshmat, 100 mm diameter self-contained solid/powder lubricated auxiliary bearing operated at 30,000 rpm. *Tribol. Trans.* **45**, 76–84 (2002)
- M.M. Khonsari, On the modeling of multi-body interaction problems in tribology. *Wear* **207**, 55–62 (1997)
- C.K.K. Lun, S.B. Savage, D.J. Jeffrey, N. Chepurniy, Kinetic theories for granular flows: Inelastic particles in couette flow and slightly inelastic particles in a general flow field. *J. Fluid Mech.* **140**, 223–256 (1984)
- K.T. McKeague, M.M. Khonsari, An analysis of powder lubricated slider bearings. *ASME J. Tribol.* **118**(1), 206–214 (1996a)
- K.T. McKeague, M.M. Khonsari, Generalized boundary interactions for powder lubricated couette flow. *ASME J. Tribol.* **118**(3), 580–588 (1996b)
- M. Pappur, M.M. Khonsari, Flow characterization and performance of a powder lubricated slider bearing. *ASME J. Tribol.* **124**, 1–10 (2002)
- M. Pappur, M.M. Khonsari, Flow characterization and performance of a powder lubricated slider bearing. *ASME J. Tribol.* **125**, 135–144 (2003)
- T. Pöschel, V. Buchholtz, Static friction phenomena in granular materials: Coulomb law versus particle geometry. *Phys. Rev. Lett.* **71**(24), 3963–3966 (1993)
- O. Reynolds, On the dilatancy of media composed of rigid particles in contact, with experimental illustrations. *Phil. Mag. Ser. 5* **20**, 469–481 (1885)
- G.W. Sawyer, J.A. Tichy, Lubrication with granular flow: Continuum theory, particle simulation, comparison with experiment. *ASME J. Tribol.* **123**, 777–784 (2001)
- H. Tsai, Y. Jeng, Characteristics of powder lubricated finite-width journal bearings: A hydrodynamic analysis. *ASME J. Tribol.* **128**, 351–356 (2006)
- M. Valenti, Upgrading jet turbine technology. *Mech. Eng.* **117**, 56–60 (1995)
- E.Y.A. Worniyoh, V.K. Jasti, C.F. Higgs, A review of dry particulate lubrication: Powder and granular materials. *ASME J. Tribol.* **129**(2), 438–449 (2007)
- C. Yu, J.A. Tichy, Granular collision lubrication: Effect of surface roughness, particle size and solid fraction. *STLE Tribol. Trans.* **39**(3), 537–546 (1996)
- C. Yu, K. Craig, J.A. Tichy, Granular collision lubrication. *J. Rheol.* **38**(4), 921–936 (1994)
- L. Zhou, M.M. Khonsari, Flow characteristics of a powder lubricant sheared between parallel plates. *ASME J. Tribol.* **122**(1), 147–155 (2000)

## Grapeseed Oil

### ► Natural Oils as Lubricants

## Graphene

### ► Solid Lubricants, Graphene

## Graphite Solid Lubrication Materials

HAI-DOU WANG

National Key Lab for Remanufacturing, Academy of Armored Forces Engineering, Beijing, People's Republic of China

### Definition

Graphite is one of three allotropic forms of the chemical element carbon; the other two are diamond and amorphous carbon. Graphite crystallizes in the hexagonal system. The carbon atoms are strongly bonded together in sheets. Because the bonds between the sheets are weak, graphite shows lower shearing strength under friction force. Thus it can be used as a solid lubricant and has become one of traditional and primary solid lubrication materials.

### Scientific Fundamentals

#### Classification

There are two main classifications of graphite: natural and synthetic.

#### Natural Graphite

Natural graphite is usually found in association with feldspars, mica, quartz, pyroxene, rutile, pyrites, and apatite. Natural graphite is subdivided into three types: flake graphite, crystalline graphite, and cryptocrystalline graphite.

#### Flake Graphite

Flake graphite occurs in metamorphic rocks uniformly distributed through the body of the ore or in concentrated lens-shaped pockets. Carbon concentrations vary between 5% and 40%. Graphite flake occurs as a scaly or lamella form in certain metamorphic rocks such as limestone, gneisses, and schists. Flake graphite is found in most parts of the world.

#### Crystalline Graphite

Crystalline graphite occurs as fissure-filled veins (Cardarelli 2008). It is typically purer than 90%. Although this type of graphite is found all over the world, it is only commercially mined in Sri Lanka by conventional shaft and surface mining techniques.

#### Cryptocrystalline Graphite

Cryptocrystalline graphite is found as minute particles in beds of mesomorphic rocks such as coal, slate, or shale deposits. The graphite content ranges from 25% to 85%

depending on the geological conditions. Cryptocrystalline graphite is the least graphitic of the natural graphites. It is extracted using conventional mining techniques and occurs primarily in Mexico, North Korea, South Korea, and Austria.

### Synthetic Graphite

Synthetic graphite can be produced from coke and pitch. It tends to be of higher purity, though not as crystalline as natural graphite. There are essentially two types of synthetic graphite. The first is electrographite, which is pure carbon produced from calcined petroleum coke and coal tar pitch in an electric furnace. The second type of synthetic graphite is produced by heating calcined petroleum pitch to 2,800°C. On the whole, synthetic graphite tends to be of a lower density, higher porosity, and higher electrical resistance. Its increased porosity makes it unsuitable for refractory applications. Synthetic graphite consists mainly of graphitic carbon that has been obtained by graphitization, heat treatment of non-graphitic carbon, or by chemical vapor deposition from hydrocarbons at temperatures above 2,100K.

### Microstructure

Graphite has a layered, planar structure, as shown in Fig. 1. In each layer, the carbon atoms are arranged in a hexagonal lattice with separation of 0.142 nm, and the distance between planes is 0.335 nm. The layers are stacked parallel to each other. The atoms within the rings are bonded covalently, while the layers are loosely bonded together by van der Waals forces. Graphite's ability to form a solid film lubricant comes from these two contrasting chemical bonds. The fact that weak Van der Waals forces govern the bonding between individual layers permits the layers to slide over one another, making it an ideal lubricant.

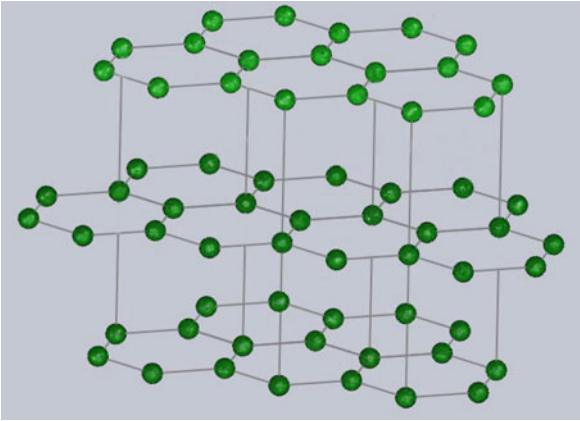
The two known forms of graphite, alpha (hexagonal) and beta (rhombohedral), have very similar physical properties except that the graphene layers stack slightly differently. The hexagonal graphite may be either flat or buckled. The alpha form can be converted to the beta form through mechanical treatment, and the beta form reverts to the alpha form when it is heated above 1,300°C.

### General Properties

The general properties of graphite are listed in Table 1. Graphite can be used as lubricant, because it has following excellent properties:

- (a) Graphite is generally grayish-black, opaque, and has a lustrous black sheen. It is soft and slippery, which gives graphite a great lubricating effect.





Graphite Solid Lubrication Materials, Fig. 1 Crystal structure of graphite

Graphite Solid Lubrication Materials, Table 1 General properties of graphite

Property	Industrial graphite
Density (kg/m <sup>3</sup> )	1,400–2,266
Melting point (°C)	3,652–3,697
Boiling point (°C)	4,250
Young's modulus (GPa)	3–12
Compressive strength (MPa)	14–42
Flexural strength (MPa)	5–21
Thermal conductivity (W/m·K)	85–350
Specific heat capacity (J/kg·K)	709
Coefficient of linear thermal expansion (10 <sup>-6</sup> /K)	1.3–3.8
Electrical resistivity (μΩ·cm)	1,385

- (b) Graphite is one of the most inert materials, and thus has a very good chemical stability; it almost doesn't react with all organic solutions or corrosive chemical reagents. So it can sustain its characteristics when contacting with solutions, oils, greases rubbers, resins, and metals.
- (c) Although graphite is a nonmetal with a very high melting temperature, it does have some characteristics of metals. It can conduct electricity just like many metals and exhibits a high thermal conductivity. Therefore, when it is used as lubricant, the heat elimination of the gliding plane cannot be impeded.

Furthermore, the electrical resistivity of graphite is higher than that of metals, and thus it has good electrical conductivity in solid lubricants.

- (d) Graphite has extremely low coefficient of linear thermal expansion, and therefore it can resist thermal shock when used as high-temperature sliding materials.
- (e) Graphite has a low adsorption of X-rays and neutrons, making it a particularly useful material in nuclear applications.

## Lubricating Properties and Influencing Factors

The lamellar structure of graphite makes it ideal to reduce friction. In the friction process, solid graphite lubricant can be transferred onto the surface of the counterpart to prevent the direct contact between the metals and the occurrence of adhesion and scuffing. When the transfer film falls off or ruptures because of fatigue or abrasion, solid graphite lubricant can be supplied continually to keep the transfer film intact and continuous. Graphite is best suited for lubrication in a regular atmosphere. Water vapor is a necessary component for graphite lubrication. The lubricity of graphite is influenced by water vapor pressure, the higher the pressure, the lower the wear rate. The adsorption of water reduces the bonding energy between the hexagonal planes of the graphite to a lower level than the adhesion energy between a substrate and the graphite. Because water vapor is a requirement for lubrication, graphite is not suitable as a solid lubricant in a dry atmosphere or vacuum. In an oxidative atmosphere, graphite is effective at high temperatures up to 450°C continuously and can withstand much higher temperature peaks.

The crystallinity, average particle size, particle shape, and impurity content of graphite can influence its lubricity (松永正久. 固体潤滑ハンドブック. 幸書房 1978):

- (a) *Influence of crystallinity.* Different types of graphite have different crystallinity. Natural scaly graphite almost crystallizes completely; its crystallization is much better than that of synthetic graphite. Generally, the higher the crystallinity, the better the lubricity.
- (b) *Influence of impurity.* In natural graphite, the impurities include iron ore, quartz, and calcite. In synthetic graphite, the impurity is iron oxide. The smaller the content of impurities in the graphite, the less the wear on the gliding plane of mechanical parts, and the longer the life of the gliding plane.
- (c) *Influence of particle size.* Whether natural or synthetic graphite, graphite powder with an average particle size



of 4~5  $\mu\text{m}$  will produce good lubrication effects and can reduce to a minimum friction and wear of a specimen. Graphite with coarse particle size can produce a grinding effect, which works against friction reduction and wear resistance. Graphite with fine particles, e.g., a particle size of 1  $\mu\text{m}$ , also leads to a decrease in friction reduction and wear resistance, which is not explained by grinding effect, because graphite with fine particles cannot reduce friction through layer slipping.

- (d) *Influence of particle shape.* The graphite used for lubrication is prepared to various powders with different particle shape, thickness, and degree of finish by different types of grinder mills such as a ball mill, hammer mill, vibrating mill, or jet pulverizer. In the graphite powders with different particle shapes, round particles exhibit better forming property, while flat particles exhibit better adhesion behavior on the lubrication surface. It is clear that the lubricating property of graphite products is affected by particle shape. The flat graphite used in oil or made for dry film has significant superior lubricating property to common graphite.

Moreover, the lubricity of graphite is also influenced by environment conditions such as gas pressure, temperature, friction velocity and friction load.

- (a) *Influence of gas pressure.* The lubricating property of graphite is greatly affected by water vapor pressure. Lubricity is also influenced by other gases such as propane, methyl alcohol, and propyl alcohol. As the gas pressure increases, the wearing rate of graphite decreases. When the particle size of graphite is very small, the influence of gases is more noticeable. In a vacuum condition, once the air, oxygen, water vapor, or gases of benzene, ammonia, ethanol, acetone, or heptane are imported, the friction coefficient will drop down. However, when the nitrogen or carbon dioxide is imported, the friction coefficient cannot be reduced.
- (b) *Influence of temperature.* It is important for graphite to be used as a high-temperature lubricant. Although the graphite has different particle sizes and crystallinity, it can be used in atmosphere at 550°C. If the oxidation inhibitor like phosphorous pentoxide is added, it can be used at 700°C. For the lubricity of graphite at high temperature, without the influence of absorbed gas in vacuum, the friction coefficient tends

to decrease at high temperature compared with that at room temperature. However, in air, the graphite loses the absorbed gas and lubricating property when the temperature just increases from room temperature to a little higher temperature. At much higher temperatures, the lubricating property appears again.

- (c) *Influence of friction velocity and friction load.* The lubricating property of graphite lubricant varies as the alteration of friction velocity and load. As the increment of friction velocity or load increases surface temperature, the absorbed gas will be lost, and thus the lubricating property will be reduced.

## Key Application

### Bulk Material

Graphite can be made into integral parts such as graphite bearings and electrical brushes by sintering or casting. Graphite bearings include carbon-graphite bearings, bronze-graphite oil-impregnated bearings, and insert bearings. Figure 2 shows a typical insert bearing.

### Dispersion

The graphite used as dispersion in solvents such as water, oil, and dissolvent has been used extensively in many areas. The solvent containing graphite dispersion supplies the graphite to lubricating surface, leaving only the graphite to work as lubricant. Therefore, first, the dispersion containing graphite should have good decentralized stability. Second, the graphite dispersion requires strong adhesion power.

### Composite Materials

Composite materials are made by combining reinforcement with matrix; this combination provides characteristics superior to either of the materials alone. Graphite can enhance the strength of the composite using amorphous carbon as adhesive agent, and it can be used as self-lubricating friction parts such as sliding bearing, sealing material, and brush.

Graphite can also serve as filler of plastic materials to enhance the load-bearing behavior and wear-resisting performance of self-lubricating plastic materials. In polyimide, the graphite exhibits more significant effects than molybdenum disulfide.

The copper-base or iron-base sintered material containing 10~30% graphite serving as sealing material or sliding bearing can be applied where the oil

cannot be fed. Copper-graphite particulate composites possess the properties of both materials, i.e., excellent thermal and electrical conductivities, and solid lubricating and small thermal expansion coefficient. Copper-graphite composites with graphite contents of 8, 15, and 20 wt% were made by powder metallurgy route using either Cu-coated graphite powders or a mixture of copper and graphite powders (Kinga and Wilman 1962). The Cu-coating process of graphite powders was carried out using an electroless coating technique. The

composites made by Cu-coated and uncoated graphite showed lower wear rates and friction coefficients than those made from pure copper.

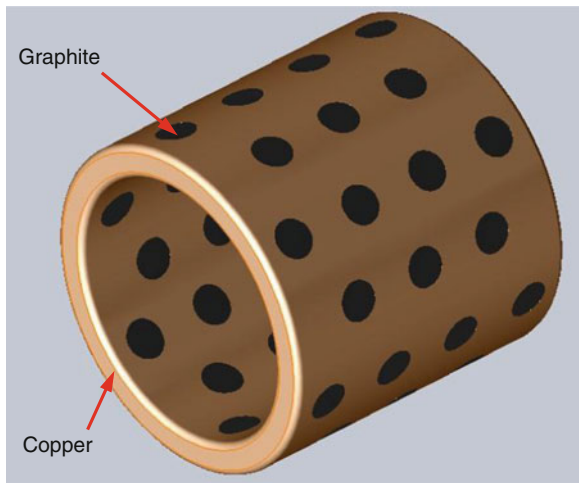
### Mixture with Oxides

For the friction surface with high temperature, if the graphite is mixed with some oxides, this combination provides good lubricity superior to either the graphite or oxides alone. For instance, at a temperature with poor lubricity, good lubricity can be sustained continually from room temperature to high temperature when the cadmium oxide is added.

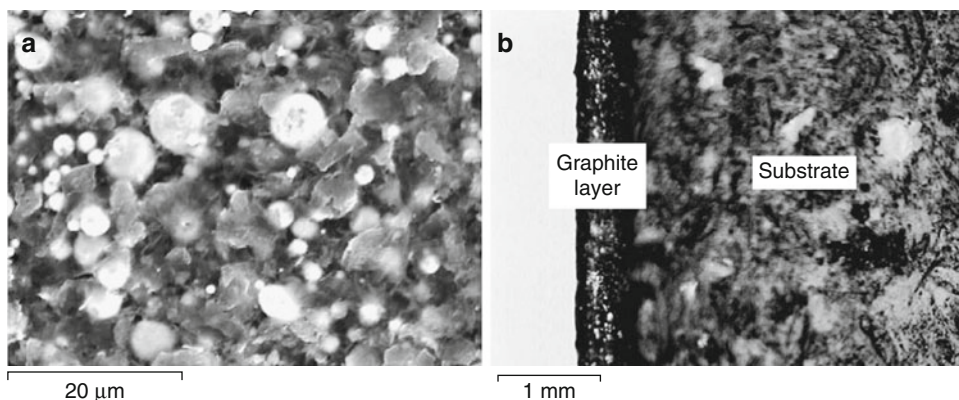
### Powders

Graphite powders can be used as an additive in lubricating oil or grease to increase their load-carrying capacity, high-temperature behavior, and to improve the boundary lubricating conditions. The graphite used as an additive to lubricating grease can effectively improve the resistance to compression and wear, and is especially suited for lubricating mechanical equipment and railway wheeltracks with high loads. Therefore, graphite is often mixed with high-viscosity mineral oil to synthesize grease. The graphite used as the additive of lubricating oil is usually compounded with engine oil and additional organic agents to prepare oil-base graphite lubricants suitable for various work conditions.

In recent years, nanoparticles have emerged as a new kind of additives in oils and greases by their special characteristics of size, shape, and other properties



Graphite Solid Lubrication Materials, Fig. 2 Copper-graphite insert bearing



Graphite Solid Lubrication Materials, Fig. 3 Morphologies of the graphite coating. (a) Surface morphology, (b) Cross-section morphology

(Xue et al. 1997; Tao et al. 1996; Qiu et al. 1999; Rapoport et al. 2003a, b). Graphite nanoparticles have been used to fabricate nano lubricants with enhanced tribological properties (Lee et al. 2009). The base lubricant used was industrial gear oil. The average size of graphite nanoparticles was 55 nm, their specific gravity was 2.26, and their concentrations were 0.1 vol% and 0.5 vol%. The tribological properties of nano lubricants were examined by a disk-on-disk tribotester. The testing results showed that the friction coefficients were remarkably reduced with

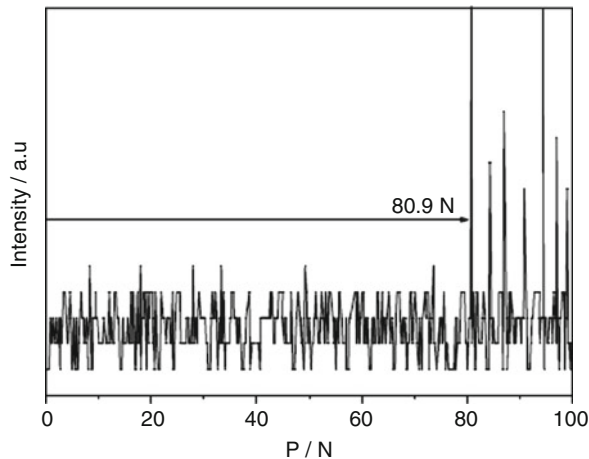
the addition of graphite nanoparticles to the base lubricant. At a normal force of 3,000 N, the friction coefficient of the nano lubricant was decreased to 24% of the raw lubricant value.

### Coatings

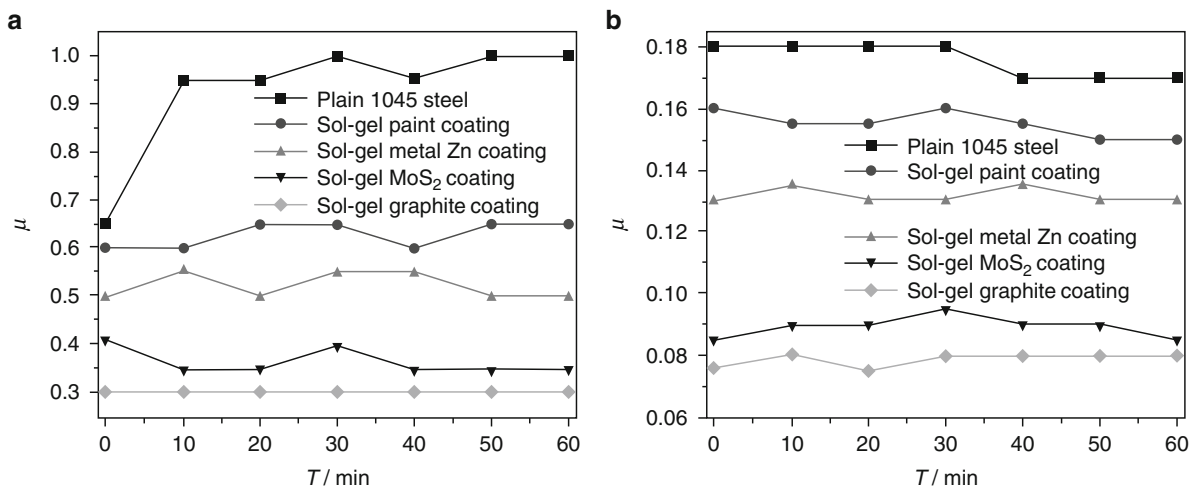
Graphite lubricant can be coated on the surface of frictional pairs as a dry film by techniques of sputtering, spraying, ion plating, brush-painting, and sol-gel (Gupta et al. 1994; Pan et al. 2010).

Graphite coating of tool steel was prepared by pressure-spraying technology (Hanada et al. 2005). A dense graphite layer of approximately 1  $\mu\text{m}$  was made on the surface of tool steel by spraying graphite powder with pressurized nitrogen gas. Friction measurements indicated that the sprayed tool steel showed excellent lubricating ability and service life compared with the case of conventional graphite coating.

A convenient and simple brush-painting technique was utilized to prepare the solid lubrication graphite coatings (Zhu 2005). The bonding between the coatings and substrate was good. To examine the influence of different graphite contents on the tribological properties of the graphite coatings, the comparison experiments were carried out on the ring-on-block tester. The testing results showed a saddle-shape variation law between the tribological properties of graphite coatings and graphite content. When the amount of graphite was up to 28 g, the tribological properties of graphite coating were the best.



Graphite Solid Lubrication Materials, Fig. 4 Bonding strength between coating and substrate



Graphite Solid Lubrication Materials, Fig. 5 Variation of friction coefficient of the different coatings (a) Friction coefficient with time under dry condition, (b) Friction coefficient with time under oil condition

The tribological properties of graphite coatings prepared by sol-gel method were also investigated (Wang et al. 2005). The colloidal solution (solvent) was a patented inorganic silicon water-based paint. The filler materials (solute) were a mixture of graphite and zinc powders with diameters of 5  $\mu\text{m}$ . The sol-gel graphite coating was about 350  $\mu\text{m}$  thick.

The morphologies of the graphite coating are shown in Fig. 3. The coating surface is compact and composed mainly of white spherical grains and gray-black netted flakes. The cross-section morphology of the graphite coating shows that the structure is compact, and the bonding between the coating and substrate is good.

The graphite coating combines well with the substrate. Figure 4 shows the bond strength between the graphite coating and substrate obtained by a scratching instrument. The first peak meant that the bonding had been destroyed and that the bond strength was up to 80 N. That is a very good value, meaning that the coating could not be easily separated from the substrate during friction.

The tribological tests under dry sliding were carried out on a ring-on-block tester model MM-200. It was found that the friction coefficient of the graphite coating was always lower than that of MoS<sub>2</sub> coating, paint coating, Zn coating, and plain 1,045 steel under both dry and oil conditions, as shown in Fig. 5.

Composite coatings have been also fabricated by laser alloying on titanium alloy Ti-6Al-4V with graphite and silicon mixed powders (Tian et al. 2006). Graphite and silicon mixed powders with a weight ratio of 1:1, average particle size of 10  $\mu\text{m}$ , blended with diluted polyvinyl alcohol solution were coated with a thickness of approximate 0.5 mm. This coating showed a lower friction coefficient and much higher sliding wear resistance compared with the as-received sample.

## Cross-References

- High-Temperature Solid Lubricating Materials
- Lubrication with a Grease
- Nanoparticle-Based Lubricant Additives

## References

- F. Cardarelli, *Materials handbook: a concise desktop reference* (Springer, London, 2008)
- B.K. Gupta, J. Janting, G. Sorensen, Friction and wear characteristics of ion-beam modified graphite coatings. *Tribol. Int.* **27**(3), 139–143 (1994)
- K. Hanada, K. Hatsukano, K. Matsuzaki, K. Umeda, M. Ishiwata, M. Tsukita, Graphite coating of tool steel by pressure spraying. *J. Mater. Process. Technol.* **164–165**, 856–861 (2005)
- J.N. Kinga, H. Wilman, The friction and wear properties, during abrasion, of compressed graphite-powder compacts and commercial graphitised carbons. *Wear* **5**(3), 213–226 (1962)
- C.-G. Lee, Y.-J. Hwang, Y.-M. Choi, J.-K. Lee, C. Choi, J.-M. Oh, A study on the tribological characteristics of graphite nano lubricants. *Int. J. Precis. Eng. Manuf.* **10**(1), 85–90 (2009)
- G. Pan, Q. Guo, J. Ding, W. Zhang, X. Wang, Tribological behaviors of graphite/epoxy two-phase composite coatings. *Tribol. Int.* **43**(8), 1318–1325 (2010)
- S. Qiu, J. Dong, G. Chen, Tribological properties of CeF<sub>3</sub> nanoparticles as additives in lubrication oils. *Wear* **230**(1), 35–38 (1999)
- L. Rapoport, V. Leshchinsky, M. Lvovsky, I. Lapsker, Y. Volovik, Superior tribological properties of powder materials with solid lubricant nanoparticles. *Wear* **255**(7), 794–800 (2003a)
- L. Rapoport, V. Leshchinsky, I. Lapsker, Y. Volovik, O. Nepomnyashchy, Tribological properties of WS<sub>2</sub> nanoparticles under mixed lubrication. *Wear* **255**(7), 785–793 (2003b)
- X. Tao, Z. Jia, X. Kang, The ball-bearing effect of diamond nano particles as an oil additive. *J. Phys.* **29**(11), 2932–2937 (1996)
- Y.S. Tian, C.Z. Chen, L.X. Chen, Q.H. Huo, Microstructures and wear properties of composite coatings produced by laser alloying of Ti-6Al-4V with graphite and silicon mixed powders. *Mater. Lett.* **60**, 109–113 (2006)
- H. Wang, B. Xu, J. Liu, D. Zhuang, The tribological properties of solid lubrication graphite coatings prepared by a sol-gel method. *Carbon* **43**(9), 2017–2020 (2005)
- Q. Xue, W. Liu, Z. Zhang, Friction and wear properties of a surface-modified TiO<sub>2</sub> nanoparticle as an additive in liquid paraffin. *Wear* **213**(1), 29–32 (1997)
- Y. Zhu, Comparison of tribological properties for graphite coatings used for remanufacturing. *J. Cent. South Univ. Technol.* **12**(2), 41–44 (2005)
- 松永正久. 固体潤滑ハンドブック. 幸書房. 1978

## Grease and Grease Life

E. R. BOOSER<sup>1</sup>, M. M. KHONSARI<sup>2</sup>

<sup>1</sup>Engineering Consultant, Vero Beach, FL, USA

<sup>2</sup>Department of Mechanical Engineering, Louisiana State University, Baton Rouge, LA, USA

## Definition

Useful life of a grease commonly reflects the drying period for its oil-thickener system. Oil content is gradually lost, predominantly by physical processes such as creep and evaporation at moderate temperatures up into the 100–125°C range; oxidation becomes more dominant at more elevated temperatures. In many ball and roller bearings, for instance, torque and noise continually increase with this dropping oil content of the grease. Failure of the greased bearing is then often encountered when about half of the original oil content in the grease is lost.

Life is also limited at subambient temperatures by the increased viscosity of the oil phase in the grease. This higher viscosity, in turn, reduces life by diminishing ability of the grease to feed into lubricated machine elements.

## Scientific Fundamentals

Selection and specifications of compositions and properties of grease requires careful consideration of guidelines for optimum performance and for estimating trouble-free life expectancy (Khonsari and Booser 2008; Potteiger 2005). While concentrating on bearing applications, the guidelines provided here should be generally useful in considering other machine elements.

Grease should normally contain the same type and viscosity of oil that would be used alone. The gelling agent is then selected for its ease in compounding the grease, for stability over a wide temperature range, water resistance, and mechanical stability under shearing action. Table 1 gives general characteristics of common mineral oil grease types.

## Oils in Greases

Petroleum mineral oils are used in more than 98% of current greases. For the low evaporation rate needed for maximum service life, combined with utility down to subzero temperatures, these oils are commonly selected in the SAE 20–30 viscosity range with a 40°C viscosity of about 100–130 cSt for long-life service in electric motors and their driven equipment. For multipurpose factory use, oils in the 150–220 cSt range are a common choice.

Higher viscosity oils up to 900 cSt at 40°C are employed for slow-speed, heavily-loaded journal bearings, gear drives, couplings, pivots, and guides operating primarily in boundary lubrication regime. They are commonly compounded with extreme-pressure additives for high contact stress at relatively low speeds. In traditional ball bearing applications, however, the lower grease mobility when incorporating these high viscosity oils commonly results in noisy operation, higher torque, shorter service life, and restricted low-temperature range.

Less viscous oils with a 40°C viscosity ranging down to about 25–50 cSt in specialty greases enable operation to lower temperatures, and freer oil feeding to cage and rolling element bearing surfaces enables higher ball bearing speeds. As a long-term effect, faster evaporation of this lower molecular weight oil significantly accelerates drying of the grease to shorten long-term grease life at elevated temperatures.

This viscosity selection is confused by the common practice in grease manufacture of blending higher and

**Grease and Grease Life, Table 1** Characteristics of common mineral oil greases

	Dropping point, °C	Temperature range, °C	Water resistance
Lithium	200	–30 to 120	Stable to 90°C
Sodium	150	–25 to 120	Poor
Calcium	90	–25 to 65	Very stable
Complex soaps	200+	–30 to 145	Stable
Bentonite	260+	–20 to 180	Very stable
Polyurea	260+	–30 to 180	Stable

**Grease and Grease Life, Table 2** Operating temperature range for synthetic greases (Khonsari and Booser 2008)

Oil Type	Common temperature range, °C
Mineral oil	–28 to 145
Diester	–56 to 125
Synthetic hydrocarbon	–40 to 145
Polyester	–46 to 160
Conventional silicone	–35 to 170
Special silicone	–73 to 230
Perfluoroalkyl polyether	–35 to 260
Polyphenyl ether	10 to 280

lower viscosity base oils to achieve the final viscosity desired. For a 110 cSt at 40°C base oil, for instance, the lighter oil used in blending might be in the 40–75 cSt range, the heavier 170–200 cSt.

Synthetic oils are used in 1–2% of current grease production where their higher cost is justified by unusual temperatures (below about –20 to –30°C or above 150–170°C) or other demands not readily met with conventional mineral oils. While synthetic hydrocarbons (polyalphaolefins, PAOs) and ester fluid-based greases enjoy the greatest volume production, almost all synthetics have been used. Table 2 compares the general operating range for synthetic greases.

## Thickeners

Common gelling agents are the fatty acid soaps of lithium, calcium, sodium, and aluminum in concentrations of 6–20 wt%. Lithium soaps are water resistant, and their original operating temperature limit in the 110–125°C



Grease and Grease Life, Table 3 NLGI consistency classification

Grade number	Worked penetration, mm <sup>-1</sup>	Approx. Yield strength, Pa <sup>a</sup>	Approx. Self-supporting height, cm	Description
000	445–475			Very fluid
00	400–430	90		Fluid
0	355–385	130	1.27	Semi-fluid
1	310–340	180	1.78	Very soft
2	265–295	300	3.05	Soft
3	220–250	560	5.59	Semi-firm
4	175–205	1,300	13.21	Firm
5	130–160		38.1	Very firm
6	85–115			Hard

<sup>a</sup>From Larsen and Bondi (1952)

range has been increased to 150–175°C by introducing a portion of low molecular weight organic acid as a complexing agent in making the soap. Since original introduction in 1942, the broad service range of lithium soaps has expanded their use to 74% of the total market.

Various non-soap, non-melting powders are finding increasing use for high-temperature service. Polyurea powders have become a common choice for long-life electric motor greases. Good mechanical stability and their non-melting character bring them into use for small to medium sizes of electric motors for operation over a wide temperature range without need for regreasing. Polyurea grease can also satisfy the high-temperature needs in furnaces and conveyors in steel mills, railroads, and marine applications up to about 175°C. Fine bentonite clay particles also find use both in mineral oil greases and in synthetic hydrocarbon greases employed in naval and nuclear power plant applications.

### Selecting Consistency

Grease stiffness is varied with thickener content and is a primary factor in grease selection. This is measured from the depth a standard cone sinks into the grease at 25°C after being worked 60 strokes by a perforated steel plate plunger (ASTM D217). This worked penetration is the basis for the consistency classification in Table 3 developed by the National Lubricating Grease Institute (NLGI). Also listed are the approximate yield strength and corresponding self-supporting height to be expected for grease in each penetration range.

The three softest grease grades, 000, 00, and 0, are semi-fluid at room temperature. Their application is limited to centralized lube systems and to multiple-row tapered roller bearings, gear boxes, and similar

applications to minimize lubricant leakage while still providing ready flow to renew grease films. Grades 5 and 6 at the other extreme are hard brick greases occasionally used in block form for such applications as sleeve-bearing boxes in kiln cars and paper machines.

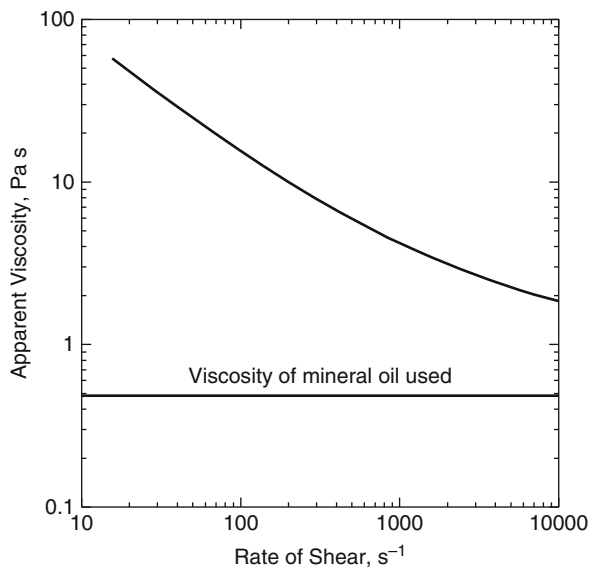
Grade 2 greases are most common. Sufficiently soft and oily to fulfill the lubrication needs of most bearings, they offer sufficient stiffness to avoid mechanical churning that would break down their gel structure. Stiffer Grade 3 greases are used in many prepacked, double-sealed ball bearings where the grease is held in close proximity to the churning action of the ball complement. This grade is also used in large bearings where the supported grease depth exceeds about 30 mm (1.2 in.) and especially when there is vibration. In these large grease cavities encountered with bearings larger than about 100-mm bore size, Grade 2 grease tends to slump into the rotating ball or roller assembly, again with churning and mechanical breakdown of the grease gel structure.

Apparent viscosity of grease at low shear rates below about 10 s<sup>-1</sup> is approximately equal to the yield value divided by the shear rate. This apparent viscosity drops rapidly as shear rate increases to about 1,000 s<sup>-1</sup> as indicated in Fig. 1 (Boner 1954). The statically stiff grease then gives an apparent viscosity of about 1.5–2 times that of its base oil in ball and roller bearings where shear rates range up to 1,000,000 s<sup>-1</sup>.

### Additives

Chemical additives similar to those used in lube oils are added to greases to improve oxidation resistance, rust protection, and wear resistance. Amines, phenolics, and sulfur oxidation inhibitors at 0.1–1.0% concentration extend both shelf life and operating life many fold.





**Grease and Grease Life, Fig. 1** Typical pattern for apparent grease viscosity as a function of shear rate (From Khonsari and Booser 2008)

Although most greases offer some inherent rust protection, additives are commonly used for added water and salt spray protection.

Various sulfur and phosphorous extreme pressure additives are widely used in multipurpose industrial greases to minimize wear at low speeds, high loads, and under shock and oscillating conditions in ball and sliding bearings in industrial plants, and in farm and construction equipment. Since these additives shorten grease oxidation life, they are not usually used in premium long-life ball bearing greases for electric motors and related units. Solid powders of molybdenum disulfide, graphite, zinc oxide, and talc are also added as fillers for extreme conditions of boundary lubrication.

## Key Applications

Grease is the lubricant of choice in millions of bearings and machine elements. Yet selection and specification of compositions and properties of grease require careful consideration of guidelines for optimum performance and for estimating trouble-free life expectancy (Khonsari and Booser 2008; Potteiger 2005). While concentrating on bearing applications, the guidelines provided here should be generally useful in considering other machine elements. Both the oil phase and the thickening agent can be selected to match needs in bearings as well as in gears, pivots, couplings, guides, and other machine elements.

**Ball and Roller Bearings.** For design simplicity, decreased sealing requirements, and low maintenance, grease is normally the first choice to lubricate small and medium-size rolling-element bearings with their very low lubrication requirements in a vast array of electric motors, household appliances, machine tools, agriculture, and construction equipment, as well as automotive and aircraft accessories.

**Journal Bearings and Sliding Contacts.** Greases find wide use to lubricate pins, bushings, and heavily-loaded sliding surfaces at low surface speeds up to 10–20 ft/min and when cooling is not a factor. Space restrictions and low maintenance favor grease rather than oil for a thicker lubricant film and less wear during boundary lubrication with shock loads, frequent stops and starts, and reversing direction in earthmoving, construction, and agricultural machinery.

Since many journal and thrust bearings need a much greater lubricant supply, grease often cannot meet their lubrication needs. As an extreme, grease cannot be considered for bearings of a large steam turbine-generator in a power house that require more than 5,000 L per minute of oil for cooling and lubrication. On a smaller scale, grease is also incapable of supplying the needs of sleeve bearings in either automobile engines or many industrial electric motors and their driven equipment.

## Typical Property and Bench Tests

While direct correlation with field performance is often difficult, the following test procedures are useful in preliminary evaluations for long grease life (Booser and Khonsari 2010).

### Dropping Point (ASTM D566 and D2265)

Dropping point is the temperature at which a drop from the grease falls from the bottom orifice in a heated test cup, commonly reflecting change from a semisolid to a liquid state. As a general rule, maximum usable temperature of grease is at least 15–30°C below this drop point.

### Evaporation Loss (ASTM D972 and D25985)

Oil loss from grease at high temperatures can lead to premature hardening and loss of lubricating properties as higher volatility (lower viscosity) portions of the oil evaporate. High temperature evaporation progressively increases as mineral oil viscosity drops below about 75 cSt at 40°C. Evaporation of 2% in 22 h at 100°C is a common specification limit for premium mineral oil greases, and ranges to 0.4% for synthetics.

### Bleeding (Federal Test Method 321.2)

Percentage of oil separating from a grease sample held in a wire cone is commonly measured after 30 h at 100°C. A 2–5% range is often desirable. No bleeding characterizes grease that fails to supply adequate lubrication to rolling-element bearings and may give noisy operation. Excess bleeding tends to lead to leakage and shortened grease life.

### Oxidation Stability (ASTM D942)

Grease is exposed to pure oxygen at 758 Pa (110 psi) at 99°C in a “bomb” for 100–500 h. Decrease in oxygen pressure gives an indication of grease life in extended storage, as in prelubricated bearings, and in high temperature service. Maximum pressure drop may be specified in the range of 20.7–172 Pa (3–25 psi) in 100 h, and 172–345 Pa (25–50 psi) in 500 h.

### Wear and Extreme Pressure Properties

The Timken test (ASTM D2509) loads a carburized steel test block against a rotating tapered roller bearing cup under a series of loads to obtain “OK” and limiting “Score” loads. Typical OK loads of 178 N (40 pounds) for multipurpose greases range up to 267 N (60 pounds) for some EP greases.

Both wear and extreme pressure properties can be evaluated in similar four-ball tests using a half-inch rotating top steel bearing ball loaded on a cradled nest of three stationary balls immersed in test grease. ASTM D2266 wear test measures scar diameter on the stationary balls after running 1 h with 40 kg load. Wear scars range from about 0.4 to 0.6 mm at 1,200 rpm and 75°C. ASTM D2596 evaluates extreme-pressure additive qualities by increasing load steps to seizure at the weld point, which ranges up to 400–500 kgf.

### Ball Bearing Grease Life

Various customized ball bearing life tests at elevated temperatures simulate automotive wheel bearings (ASTM D3527), electric motor bearings, military units, and other equipment. ASTM D3336 runs greased 204 (20-mm bore) ball bearings at 10,000 rpm and elevated temperature either to failure or to a specified running time. Results at these high speeds and high temperatures can be misleading: insufficient oil bleeding or excessive channeling could result in early failure of a NLGI Grade 3 long-life grease, which would otherwise be ideal for sealing and for low-speed bearings.

ASTM D1741 uses 306 (30-mm bore) ball bearings at 3,600 rpm to evaluate grease life at 125°C for electric motor bearings and similar industrial uses. Life of 10,000 h can be expected with premium Grade 2 lithium and polyurea greases.

### Operating Grease Life Factors

During extended operation, oil content of grease is slowly lost by creep, evaporation, and oxidation. As the grease dries, it stiffens and darkens with loss of its ability to lubricate.

To obtain the longest possible operating life, an oxidation inhibitor is essential in premium and multipurpose greases, along with their usual rust inhibitor. Common anti-wear and extreme pressure additives in multipurpose greases, while enabling superior performance in boundary lubrication, greatly reduce oxidation life and correspondingly require more frequent relubrication in usual ball and roller bearing service.

Before adopting grease for general use, trials in actual applications are desirable. These trials should be conducted under the most severe conditions expected: high temperature, water spray, over-packed bearing, and mixing with existing grease. High power loss, high temperature rise, leakage, wear, or noisy operation would indicate inadequate performance. During continued running, periodic checks for drying of the grease, higher acidity, or darkening color would raise questions as to long-term life.

### Grease Life in Ball and Roller Bearings

With premium ball bearing greases, grease life involving 10% failures at mild operating conditions and temperatures up to 70°C is commonly about 40,000–45,000 h. During this period, about half of the initial oil content of the grease is commonly lost, and grease replacement is needed as bearing friction and noise slowly rise as failure approaches. For maximum reliability, the regreasing interval should be reduced to about half of this 10% failure time.

For bearing temperatures above 70°C, grease life  $L$  in hours tends to drop by a factor of 1.5 for each 10°C temperature rise. This corresponds to the 2450 activation energy term in the Arrhenius reaction rate Eq. 1 and is reflected in Fig. 2 (Booser and Khonsari 2007)

$$\log L = -2.60 + 2450/(273 + T) - (9.6 \times 10^{-7})k_f DN \quad (1)$$

where  $L$  is the grease life in hours to reach 10% failures,  $T$  is temperature in °C,  $DN$  is the product of bore diameter (mm)  $\times$  shaft speed (rpm), and  $k_f$  is a constant provided in Table 4.

This 1.5 per 10°C factor reflects predominance of physical factors such as oil creep and evaporation as primary life limits. Above about 150–160°C this factor rises to 2.0 per 10°C with the corresponding activation energy

**Grease and Grease Life, Table 4** Speed Factor  $k_f$  for grease life reduction in (1)

Bearing Type	Factor $k_f$
Deep-groove ball bearing	
Single row	0.9–1.1
Double row	1.5
Angular-contact ball bearing	
Single row	1.6
Double row	2
Spindle bearing	
$\alpha = 15^\circ$	0.75
$\alpha = 25^\circ$	0.9
Four-point bearing	1.6
Self-aligning ball bearing	1.3–1.6
Thrust ball bearing	5–6
Angular-contact thrust ball bearing, double row	1.4
Cylindrical roller bearing	
Single row	1.8–2.3
Double row	2
Full complement	25
Cylindrical roller thrust bearing	90
Needle roller bearing	3.5
Tapered roller bearing	10

Source: FAG (1995)

term rising to 6,000 to reflect the growing predominance of oxidation as the life-controlling factor in the relation (Booser and Khonsari 2010; Rhee 2008):

$$\log L = -10.79 + 6000/(273 + T) - (9.6 \times 10^{-7})k_f DN \quad (2)$$

At the other extreme, grease life commonly begins to drop as bearing temperature ranges below about 30–40°C. At low temperatures, grease with its higher viscosity oil stiffens and feeds less readily for lubrication of bearing surfaces (Smith and Wilson 1980; Booser and Khonsari 2010).

Life at any bearing temperature also drops to about half of that in Fig. 2 as the bearing DN speed factor rises to the 250,000–300,000 range. Above this “limiting speed” for grease lubrication, centrifugal throw-off of grease from bearing surfaces and need for quicker replenishment of lubricant films on contact surfaces leads to erratic performance and short life.

Further grease life reduction given by the final term of (1) and (2) depends on the type of ball or roller bearing

with their  $k_f$  values of Table 4. Higher  $k_f$  values reflect additional grease feeding needs to satisfy larger and more demanding rolling element geometries. While grease life also decreases with bearing load, the effect is only minor for the relatively light ball bearing loads in electric motors and most other long-life industrial equipment (Booser 1974).

**Example.** What grease life is to be expected in a 215 size deep-groove ball bearing (with 75-mm bore diameter  $D$ ) at 85°C in a 3,600 rpm electric motor?

Applying (1) for premium greases, and with  $k_f = 1.0$  from Table 4:

$$\begin{aligned} \log L &= -2.60 + 2450/(273 + T) + (9.6 \times 10^{-7})(1.0) \\ &\quad (75 \times 3600) = 3.985 \\ L &= 9660 \text{ h} \end{aligned}$$

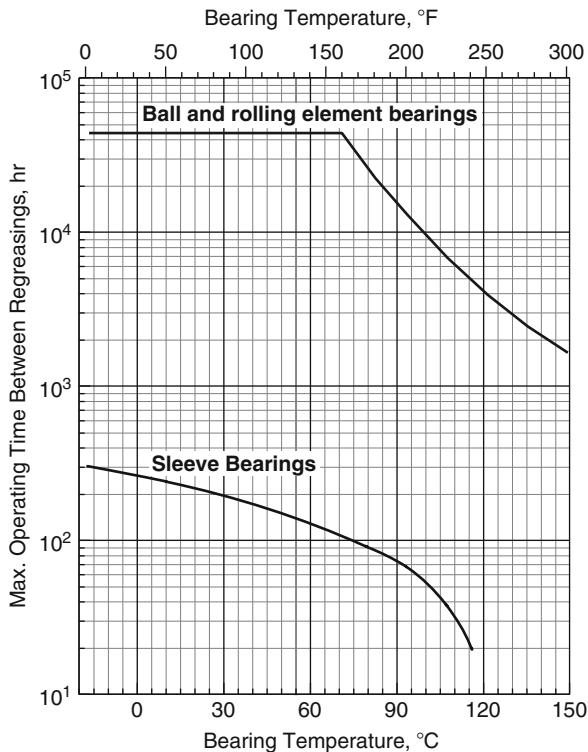
Neglecting the final speed factor term  $(9.6 \times 10^{-7})k_f DN$  would increase the expected life  $L$  to 17,500 h.

## Grease Life in Journal Bearings and Other Sliding Contacts

Grease lubrication is used for pins, bushings, and sliding contacts, and for journal bearings experiencing shock loads, frequent starts and stops, or reversing direction. In such cases, grease provides a thicker lubricant film than oil and protects against wear during the boundary lubrication commonly involved.

Typically, a Grade 2 grease is employed with a base oil viscosity in the 150–220 cSt range at 40°C. Higher viscosity base oils and extreme pressure and solid additives are brought into use for low speeds, high loads, and high temperature.

Grease used in a plain bearing must be replenished periodically. Frequency needed for this regreasing depends on the lubricating efficiency of the grease, its thermal stability, and rate at which wear debris is produced. Bearing temperature has the strongest influence on this grease life. As with ball bearings, higher temperature increases the rate of oil loss from the grease structure by oxidation, creep, and evaporation. With this drying, grease hardens and loses its ability to distribute itself in the bearing and replenish boundary films. Figure 2 gives recommended time of continuous operation for applications such as sleeve bearings in farm and construction machinery as a function of bearing temperature (Glaeser and DuFrane 1978; Booser et al. 2010). Some experimental results for grease-lubricated journal bearings operating under oscillatory conditions are available from Lu and Khonsari (2007a, b).



**Grease and Grease Life, Fig. 2** Guidelines for bearing regreasing intervals with mild operating conditions (Adapted from Khonsari and Booser 2007)

Another concern with heavily loaded construction machinery is their oscillatory nature of operation—a condition that can be more severe than the continuous, unidirectional speeds typical of most machines. Here, galling or scuffing failure can be a concern. Experiments show that friction coefficient can suddenly rise after a number of cycles with little warning. The sudden rise can be induced by many factors, including wear of the protective layer, introduction of wear particles in the contact, or a sudden change from external sources such as a shock. Oscillatory speeds under very low amplitudes and high frequency can also restrict replenishment of grease into the contact, causing fretting failure as well as false brinelling.

While grease lubrication is generally impractical above 3–6 m/min (10–20 ft/min), this limit is extended with continuous feed systems applying semi-fluid NLGI 00, 0, or 1 grades. As a rule of thumb, sufficient grease should be fed to coat the contact area with a 0.07-mm-thick lubricant layer every hour.

## Cross-References

- [Lubrication with a Grease](#)
- [Oil Life](#)
- [Rolling Bearing Lubricants](#)

## References

- C.J. Boner, *Manufacture and Application of Lubricating Greases* (Reinhold Publishing, New York, 1954)
- E.R. Booser, Grease life forecast for ball bearings. *Lubr. Eng.* **30**, 536–541 (1974)
- E.R. Booser, Life of oils and greases, in *Tribology Data Handbook* (CRC Press, Boca Raton, 1997), pp. 1018–1028
- E.R. Booser, M.M. Khonsari, Grease life in ball bearings: the effect of temperature. *Tribol. Lubr. Technol.* **66**(10), 36–44 (2010)
- E.R. Booser, M.M. Khonsari, Systematically selecting the best grease for long life. *Mach. Lubr.* Jan–Feb 2007, pp. 19–25
- E.R. Booser, M.M. Khonsari, R. Miller, Grease keeps sleeve bearings lubricated. *Mach. Design*, Oct 2010, pp. 54–59
- FAG Bearing Co. FAG rolling bearings, Catalog WL41 520 ED. FAG Bearings Corp., Danbury, (1995)
- W.A. Glaeser, K.F. Dufrane, New design methods for boundary-lubricated sleeve bearings. *Mach. Design* **6**, 207–213 (1978)
- M.M. Khonsari, E.R. Booser, *Applied Tribology-Bearing Design and Lubrication*, 2nd edn. (Wiley, Chichester, 2008)
- R.G. Larsen, A. Bondi, Lubricants and lubrication. *Encyclopedia of Chemical Technology*, vol 8, Interscience Encyclopedia, (1952), pp. 495–540
- X. Lu, M.M. Khonsari, An experimental analysis of grease-lubricated journal bearings. *ASME J. Tribol.* **129**, 640–646 (2007a)
- X. Lu, M.M. Khonsari, An experimental study of grease-lubricated journal bearings undergoing oscillatory motion. *ASME J. Tribol.* **129**, 84–90 (2007b)
- J. Potteiger, Step-by-step grease selection. *Mach. Lubr.* Sept–Oct 2005, pp. 54–56
- I.-S. Rhee, Decomposition kinetics of greases by thermal analysis. *NLGI Spokesman*, May 2008, pp. 34–42
- R.L. Smith, D.S. Wilson, Reliability of grease-packed ball bearings for fractional horsepower motors. *Lubr. Eng.* **36**, 411–416 (1980)

## Grease Composition

- [Lubricating Grease](#)

## Grease for Rolling Bearing Lubrication

- [Rolling Bearing Lubricants](#)

## Grease Formulation

- [Lubricating Grease](#)

## Grease Gelling Agents

### ► Grease Thickeners

## Grease Hardening Compounds

### ► Grease Thickeners

## Grease Lubrication Applications

### ► Rolling Bearing Lubrication

## Grease Thickeners

JOSEPH P. KAPERICK  
R&D Department, Afton Chemical Corporation,  
Richmond, VA, USA

### Synonyms

Grease gelling agents; Grease hardening compounds

### Definition

Grease thickeners are chemical compounds that are used as one of the key components in formulation of greases. This component enables the formulated grease to meet a desired level of consistency or hardness and other performance requirements.

### Scientific Fundamentals

#### Background

The art of making grease is a relatively new science. Most greases as we know them today were not commercialized until the early to mid-twentieth century.

Grease is typically defined as a dispersion of a thickening agent in a liquid lubricant. Many traditional grease thickeners are formed by reacting an acid with a base to form a soap. This reaction is commonly referred to as *saponification* and the thickeners formed in this manner are known as soaps, or soap thickeners. A series of intertwined fibers forms a matrix that holds the oil,

much in the same manner that a sponge absorbs water. Some non-soap greases trap the oil in a different type of matrix.

Greases can vary from a semi-fluid consistency that will pour to a block grease that can be cut into a brick-like shape and applied to the application by hand. The hardness of grease is largely due to the amount of thickener in the product. In general, the more thickener that is present, the harder the grease will be.

Most traditional grease compositions are as follows:

Oil	70–90 wt%
Thickener	2–20 wt%
Additives	0–15 wt%

Each of these components can have a significant impact on the performance of the grease.

Grease offers some advantages over fluid lubricants. These include:

- The grease thickener holds the lubricant in place.
- Grease helps to seal out contaminants.
- Grease stays in place in applications with intermittent operation.
- Grease reduces leakage.
- Grease can be used to suspend solid additives.

Fluids can also have certain innate benefits, such as better heat dissipation, improved flushing of contaminants from the contact zone, filterability, and potentially lower cost.

### Base Stocks Commonly Used in Grease

The type of oil or oil blend used in grease depends on the desired performance and/or proposed application for the product. Oils commonly used in the grease industry are naphthenic (cyclic hydrocarbons with single C–C bonds), paraffinic (chain hydrocarbons with single C–C bonds), synthetic (synthesized oils), and vegetable (seed oils). Oils that have single bonds between the carbon atoms are referred to as *saturated* and oils that have double bonds between the carbon atoms are referred to as *unsaturated*. Unsaturated oils are generally less thermally stable and tend to be more easily oxidized. Saturated oils are typically desirable for use in grease.

**Naphthenic Oils:** Naphthenic oils generally have lower pour points than paraffinic oils, assuming that the oils have similar viscosities. This can be an advantage in some low-temperature applications. Naphthenic oils typically have good solubility, which helps this type of oil to incorporate well into the grease structure and mix well with

most additives. The good solubility of this oil can promote good thickening (yields) and can, in some cases, contribute to lower oil separation characteristics.

**Paraffinic Oils:** Paraffinic oils generally have better thermal stability and lower evaporation rates than comparable naphthenic base oils. Paraffinic oils typically have a higher viscosity index than comparable naphthenic oils.

**Synthetic Oils:** Many synthetic oils have excellent thermal stability, low pour points, and good low-temperature characteristics. They are generally more expensive than naphthenic and paraffinic base oils. Due to a lower solubility, some synthetic oils can decrease grease yields. Many synthetic oils have a high viscosity index and low evaporation rates.

**Vegetable Oils:** Some vegetable oils exhibit good biodegradability. They are a renewable source of base oils.

Blends of oils are commonly used in order to achieve the desired viscosity and performance profile for a specific application.

### Significance of Oil Specifications for Grease

Specification sheets for oils commonly include the following characteristics. Their significance in grease manufacturing is as follows:

**Viscosity:** The viscosity is a measure of the oil's tendency to flow at specified temperatures. Oils with higher viscosities do not flow as readily at the same temperature as oils with lower viscosities. The desired application for the grease determines the optimum viscosity profile.

**Viscosity Index:** This is a measure of the degree to which the viscosity will change over a defined temperature range. It is sometimes desirable for a grease to have similar flow characteristics at various temperature intervals. The higher the viscosity index (VI), the lower the deviation in the viscosity as the temperature changes.

**Pour Point:** The pour point is the temperature at which the oil ceases to flow under defined conditions. Oils with lower pour points are sometimes preferred for low-temperature grease applications.

**Flash Point:** The flash point is the lowest temperature at which the oil vapors will ignite when exposed to a flame under defined conditions. The flash point of an oil should be high enough to be safe in the grease manufacturing process and in the intended application.

**Color:** The color is often determined by comparing the color of the oil to the color of various standards. Since a significant amount of oil is present in grease, the color of the base oil can have an impact on the color of the grease.

**Aniline Point:** In general, better grease yields can be obtained with oils that have lower aniline points, which

indicate a higher degree of solvency. Better elastomer compatibility can sometimes be obtained by using oils with higher aniline points.

**Odor:** The importance of this characteristic is dependent on the application.

### Additive Selection

Additives can be added to enhance various performance traits of the finished product. Rust inhibitors, copper corrosion inhibitors, oxidation inhibitors, antiwear/extreme pressure additives, polymers, and dyes are common additives used in grease products. The amount and combination of additives depends on the desired performance of the product.

### Manufacturing

There are numerous ways that grease can be manufactured. Greases can be reacted in open kettles, pressurized kettles, or in a continuous manufacturing system. The heat ramps, the time that the product is held at various temperatures, the mixing system, and the cooling rates can be critical in determining the final consistency of the finished product.

### Key Applications

There are a wide variety of grease types. Each grease type has unique strengths and weaknesses. Greases can be formulated and manufactured in a wide variety of ways. Therefore, exceptions to the following general descriptions are possible.

### Soap-Thickened Greases

#### Hydrated Calcium Grease

Hydrated calcium greases are typically made by reacting lime (calcium hydroxide) with either stearic acid or tallow, which is a mixture of many acid types (primarily palmitic, stearic, and oleic). The source of tallow is generally beef or mutton. This thickener is sometimes referred to as "cup grease" due to the early application method of placing this type of grease in a cup to which pressure was applied to dispense the grease. Hydrated calcium grease is also sometimes referred to as calcium stearate grease. [Table 1](#) lists the strengths and weaknesses of hydrated calcium grease.

Antioxidants are sometimes incorporated in cup greases. Hydrated calcium soaps made with tallow can discolor (turn orange/brown) if additives are not present to inhibit the oxidation process. Extreme pressure and antiwear additives are less common since other grease types are often selected for severe service applications.



**Grease Thickeners, Table 1** Hydrated calcium grease: typical strengths and weaknesses

Strengths	Weaknesses
Good water resistance	≈200°F dropping point
Good pumpability	Cannot be used above the boiling point of water
Relatively low cost	Poor oxidation resistance when made with tallow

Due to their good water resistance, low cost, and good pumpability, this grease type is sometimes used for rail-road curve applications. Graphite is often incorporated in the grease for this application.

Hydrated calcium greases can lose their thickening structure if heated above the boiling point of water, thus limiting their use in higher temperature applications.

### Sodium Grease

Sodium greases are commonly made by reacting caustic (sodium hydroxide) with a fat or fatty acid. A wide variety of fats or fatty acids can be used in the manufacturing of sodium greases. Tallow is sometimes chosen as a reactant due to the lower cost of this material. Stearic acid can also be used. The properties of this grease can vary somewhat depending on the type of fatty acid used to make the grease. [Table 2](#) lists the strengths and weaknesses of sodium grease.

Antioxidants are commonly incorporated into sodium greases, particularly if the grease is made with tallow. Rust inhibitors are generally not required. Sodium grease can have a tendency to absorb water and pull the moisture away from the metal surface, thus reducing corrosion on the metal surface. Extreme pressure and antiwear additives can be added if needed for the desired application. The texture and fiber length of sodium greases can vary depending on the oil and fat/fatty acid selection. Many sodium greases have a very fibrous texture, which helps to hold the grease in place. The long fibers, however, can reduce the pumpability of this grease type.

Sodium greases are sometimes used in applications requiring good adhesion and corrosion resistance. Other grease types are often preferred for applications where water is present.

### Anhydrous Calcium Grease

Anhydrous calcium grease is typically formed by reacting lime (calcium hydroxide) with 12-hydroxystearic acid.

**Grease Thickeners, Table 2** Sodium soap: typical strengths and weaknesses

Strengths	Weaknesses
Fibrous texture can be achieved	Can be soluble in water
Natural rust protection	Long fibers can inhibit pumpability
Long fiber version tends to stay in place well	Poor oxidation resistance when made with tallow

**Grease Thickeners, Table 3** Anhydrous calcium: typical strengths and weaknesses

Strengths	Weaknesses
H-1 (incidental food contact) potential	≈280–310°F dropping point (adequate for many applications, but not for some high-temperature applications)
Good pumpability	
Good shear stability	
Translucent appearance potential	
Good water resistance	

This grease is sometimes referred to as calcium 12-hydroxystearate grease. [Table 3](#) lists the strengths and weaknesses of anhydrous calcium grease.

Anhydrous calcium grease can be used in H-1 (incidental contact) food grade applications if the oils and additives are chosen appropriately. Anhydrous greases typically pump well, maintain a good consistency in storage, are relatively shear stable, have good resistance to water, and respond well to additives. This thickener functions well in a wide variety of multi-purpose applications.

### Lithium Soap Grease

Lithium soap greases are generally made by reacting lithium hydroxide monohydrate with one of a variety of long-chain fatty acids to form a lithium soap. Stearic acid is typically reacted to make lithium stearate grease, and the most common form of stearic acid, 12-hydroxystearic acid (derived from castor oil) is typically reacted to make a lithium 12-hydroxystearate grease. The desired price range and performance characteristics determine the type of acids or acid combinations that should be selected for this thickener. [Table 4](#) lists the strengths and weaknesses of lithium soap grease.

**Grease Thickeners, Table 4** Lithium soap: typical strengths and weaknesses

Strengths	Weaknesses
Good flow properties	Not appropriate for extreme high-temperature applications
Good storage characteristics	Moderate cost
Smooth texture	
Good shear stability	
≈350–400°F dropping point	
Good water resistance	
Low oil separation	

Lithium soaps account for a significant percentage of the grease market. They have good compatibility with many other thickener types. Lithium soaps respond well to additives and typically have good storage stability, water resistance, oil separation, and pumpability. Lithium soap greases are commonly used in a wide variety of multi-purpose applications.

### Soap Mixtures

Soap mixtures are sometimes utilized to improve certain performance characteristics. This is done by combining a thickener that has poor performance in a particular area with a compatible thickener that has good performance in the particular area trait. For example, the water resistance of a lithium grease can sometimes be improved by formulating a lithium/calcium grease. The blend can have a lower cost than that of a pure lithium soap. The dropping point of the blend, however, can be lower than that of a pure lithium soap. The water resistance of sodium grease can sometimes be improved by preparing a sodium/calcium grease. Numerous other blend possibilities exist. Blends should be considered when there is a performance/cost advantage for the desired application.

## Soap Complex Thickened Greases

### Aluminum Complex Grease

Aluminum complex greases can be formed by reacting an aluminum isopropoxide with benzoic acid and stearic acid to form an aluminum benzoate stearate complex. The structure of aluminum complex grease is somewhat different from the rope-like structure of lithium and

**Grease Thickeners, Table 5** Aluminum complex: typical strengths and weaknesses

Strengths	Weaknesses
Good water resistance	Can false set in storage
H-1 potential	Can have higher fretting wear than some thickener types
Transparent appearance potential	
≈500°F + dropping point	
Thermo-rheological reversibility	
Good pumpability	

**Grease Thickeners, Table 6** Lithium complex: typical strengths and weaknesses

Strengths	Weaknesses
Good compatibility with many other thickener types	Can cost more than some other thickener types
Good storage stability	
Good pumpability	
Good water resistance	
≈500°F + dropping point	

lithium complex greases. [Table 5](#) lists the strengths and weaknesses of aluminum complex grease.

Aluminum complex greases are often used in H-1 food-grade applications. They are also sometimes used in the steel industry due to their high dropping points, good pumpability, reversibility (ability to go to a full melt and then reform into a thickener when cooled), and their good water resistance. Aluminum complex grease can sometimes false set or harden when left undisturbed over time, but will typically revert to the original hardness as soon as the grease is disturbed or mixed.

### Lithium Complex Grease

Lithium complex greases are commonly made by reacting lithium hydroxide monohydrate with a dibasic acid and 12-hydroxystearic acid. The resulting soap typically has a very fibrous network that provides good performance in a wide variety of applications including higher temperature applications for which simple lithium greases are not typically recommended. [Table 6](#) lists the strengths and weaknesses of lithium complex grease.

Lithium complex greases are among the more popular high-temperature grease types because their dropping points are relatively high. They tend to have good storage stability, are relatively easy to pump, generally respond well to additives, and provide good water resistance. Lithium complex greases are used for a wide variety of mining, automotive, agricultural, and industrial applications. They are compatible with many other traditional grease types.

### Calcium Complex

Calcium complex greases can be formed by reacting lime (calcium hydroxide) with acetic acid, 12-hydroxystearic acid, and/or a mixture of other fatty acids.

This grease tends to have good water resistance and a relatively high dropping point ( $\approx 500^\circ\text{F}+$ ). It can have some inherent extreme pressure protection and has the potential to be used in H-1 (incidental contact) food-grade applications. This thickener can age harden and compatibility may be an issue with some other thickener types.

## Non-soap Thickened Greases

### Clay Grease

Clay greases can be formed by dispersing the platelets in an activated clay or bentonite. This distribution traps the oil in the thickener system. Clay greases can be made without the addition of the heat that is typically required for the formation of traditional soap greases. Table 7 lists the strengths and weaknesses of clay grease.

Clay greases are often referred to as non-melt greases due to the absence of a true dropping point. This trait is the reason many clay greases are used in high-temperature applications such as autoclaves, oven conveyors, and asphalt pavers. Care must be used when mixing clay greases with other thickener types as incompatibility can cause clay greases to soften when mixed with some greases or additives.

### Polyurea Grease

Polyurea is a non-soap thickener commonly formed by reacting an amine with an isocyanate. Polyurea greases are made in a variety of ways, as different versions of this thickener will have a wide range of desirable performance traits. Table 8 lists the strengths and weaknesses of polyurea grease.

Due to the absence of a metal catalyst, polyurea thickeners tend to have good oxidation resistance. For this reason, they are sometimes used in fill-for-life

**Grease Thickeners, Table 7** Clay grease: typical strengths and weaknesses

Strengths	Weaknesses
Can have good high-temperature performance	Can separate oil in storage
Non-melt (very high or no dropping point)	Can be incompatible with some additives and thickener types (e.g., clay structure can be disrupted by polar additives such as rust inhibitors)
H-1 potential	

**Grease Thickeners, Table 8** Polyurea: typical strengths and weaknesses

Strengths	Weaknesses
Oxidation resistance is typically excellent	Rheological sensitivity (some versions may soften when worked)
Good compatibility with some elastomer types	Some versions may be incompatible with more thickener types
Can have a low fretting wear	
$\approx 500^\circ\text{F} +$ dropping point	
H-1 potential	

applications. Polyurea greases can have good compatibility with some elastomer types and are sometimes used in constant velocity joint (CVJ) applications. The compatibility and the worked stability of this grease type depend largely on the reactants used to make the grease. Some versions of this thickener type have good compatibility and worked stability, but caution should be used since this is not true of all variations.

### Calcium Sulfonate

Calcium sulfonate greases can be made in a wide variety of ways. The thickener generally consists of micelles instead of fibrous strands. It is difficult to make general statements regarding compatibility of this grease with other grease types due to the diversity of the ways in which this thickener can be made. Calcium sulfonate greases tend to have high dropping points [ $\approx 260^\circ\text{C} + (\approx 500^\circ\text{F}+)$ ] as well as providing outstanding extreme pressure protection and corrosion resistance.

The thickener content of calcium sulfonate grease is typically much higher than that of other traditional grease types. The thickener content can have an impact on the cost of a product since the thickener is often an expensive component of a grease.

### Silica Grease

Silica grease can be an attractive clear grease depending on the components used to make the product. It can be made with a fumed silica and, like clay, can be made without the addition of heat. Silica greases have an H-1 potential and a relatively high ( $\approx 500^\circ\text{F}$ ) dropping point. They can have varying degrees of water resistance, depending on the type of silica used to make the product.

### Other Grease Types

There are many additional grease types. Some are used in specialized applications or extreme environments, while others can be used in a variety of multi-purpose applications. Solid additives such as PTFE and graphite can sometimes be used to thicken grease. Other grease types include titanium complex, barium soap, barium complex, and sodium complex.

The best grease for a particular application often depends on the thickener type, oil selection, additive combination, and manufacturing procedure. A review of the product specification sheet is often helpful in making the most appropriate selection for your application.

### Cross-References

- [Detergents](#)
- [Grease Composition](#)
- [Lubricant Viscosity](#)
- [Lubricating Grease Base Stocks](#)
- [Lubrication with a Grease](#)
- [Rolling Bearing Lubricants](#)

### References

- C.J. Boner, *Manufacture and Application of Lubricating Greases* (Hafner Publishing Company, New York, 1966)
- C.J. Boner, *Modern Lubricating Greases* (Scientific Publications (G.B.), Broseley, 1976)
- R.L. Frye, *An Introduction to Lubricating Greases* (Amoco Oil Company, Naperville)
- W.C. Howell Jr., *Soap Greases-Simple and Mixed Base* (Exxon Research and Engineering Company, Linden, 1979)
- H.G. Kirschenbauer, *Fats and Oils* (Reinhold Publishing Corporation, New York, 1960)
- F. Litt (ed.), *Starting from Scratch: Tribology Basics* (Society of Tribologists and Lubricating Engineers, Park Ridge)
- T.G. Musilli, *An Introduction to Lubricating Greases* (Battenfeld Grease and Oil Corporation, New York)

NLGI, *NLGI Lubricating Grease Guide* (NLGI, Kansas City, 1987)

R.A. Storer (ed.), *Annual Book of ASTM Standards*, vol. 3 (American Society for Testing and Materials, Philadelphia, 1995)

R.O. Walker (ed.), *Official and Tentative Methods of the American Oil Chemists' Society*, vol. 2 (American Oil Chemists' Society, Champaign, 1982)

T.E. Wood, Analytical Methods, Evaluation Techniques, and Regulatory Requirements, in *Soap Technology for the 1990's*, ed. by L. Spitz (American Oil Chemists' Society, Champaign, 1991)

## Green Textiles

- [Sustainable Technology for Tribological Textiles](#)

## Green's Functions in Contact Mechanics

LEON M. KEER

Department of Mechanical Engineering, Northwestern University, Evanston, IL, USA

### Definitions

Hills and colleagues (1996) define a Green's function as "the response of a system to a standard input or, in elasticity theory, the stress/strains induced in a body in response to, for example, a point force." Results derived from Fourier transforms are also presented in Mura (1993).

### Scientific Fundamentals

#### History

Problems in elasticity theory can be solved using potentials. In particular, the definition of a Green's function given above can be used mathematically to derive solutions to point load problems, either within the elastic body or on its surface. Such problems were treated by Boussinesq (1885) and later reproduced by Green and Zerna (1968). This history and excellent examples are given by Barber (2010). Using the potentials, Green's functions for some common concentrated loadings can be derived. Since the objective of this section is not to derive the Green's functions, some of them will be written out in full.

### Formulae for Common Examples

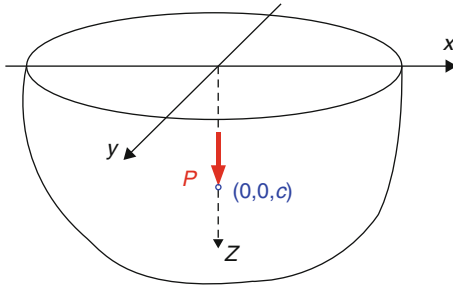
The examples used here will be from three-dimensional elasticity and will be given in terms of the fundamental solution by Mindlin (1936), who solved the problem of a point in the interior of a semi-infinite solid. This remarkable solution gave solutions to the problems of Kelvin (point in an infinite solid), Boussinesq (vertical point load on the surface of a semi-infinite solid), and Cerruti (horizontal point load on the surface of a semi-infinite solid) as special cases. It should be also noted that the solution of a point force in the interior of one of the two joined semi-infinite solids, Mindlin's solution, is given next. The solution is reproduced in full due to its utility for researchers.

### Mindlin Problem

$$R_1 = \sqrt{x^2 + y^2 + (z - c)^2}$$

$$R_2 = \sqrt{x^2 + y^2 + (z + c)^2}$$

### Vertical Load



$$u_x = \frac{Px}{16\pi\mu(1-\nu)} \left[ \frac{z-c}{R_1^3} + (3-4\nu) \frac{z-c}{R_2^3} - \frac{4(1-\nu)(1-2\nu)}{R_2(R_2+z+c)} + \frac{6cz(z+c)}{R_2^5} \right]$$

$$u_y = \frac{Py}{16\pi\mu(1-\nu)} \left[ \frac{z-c}{R_1^3} + (3-4\nu) \frac{z-c}{R_2^3} - \frac{4(1-\nu)(1-2\nu)}{R_2(R_2+z+c)} + \frac{6cz(z+c)}{R_2^5} \right]$$

$$u_z = \frac{P}{16\pi\mu(1-\nu)} \left[ \frac{3-4\nu}{R_1} + \frac{(3-4\nu)^2 - 4(1-\nu)(1-2\nu)}{R_2} + \frac{(z-c)^2}{R_1^3} + \frac{(3-4\nu)(z+c)^2 - 2cz}{R_2^3} + \frac{6cz(z+c)^2}{R_2^5} \right]$$

$$\tau_{xx} = \frac{P}{8\pi(1-\nu)} \left[ \frac{(1-2\nu)(z-c)}{R_1^3} - \frac{3x^2(z-c)}{R_1^5} + \frac{(1-2\nu)[3(z-c) - 4\nu(z+c)]}{R_2^3} - \frac{3(3-4\nu)x^2(z-c) - 6c(z+c)[(1-2\nu)z - 2\nu c]}{R_2^5} - \frac{30cx^2z(z+c)}{R_2^7} - \frac{4(1-\nu)(1-2\nu)}{R_2(R_2+z+c)} \times \left( 1 - \frac{x^2}{R_2(R_2+z+c)} - \frac{x^2}{R_2^2} \right) \right]$$

$$\tau_{yy} = \frac{P}{8\pi(1-\nu)} \left[ \frac{(1-2\nu)(z-c)}{R_1^3} - \frac{3y^2(z-c)}{R_1^5} + \frac{(1-2\nu)[3(z-c) - 4\nu(z+c)]}{R_2^3} - \frac{3(3-4\nu)y^2(z-c) - 6c(z+c)[(1-2\nu)z - 2\nu c]}{R_2^5} - \frac{30cy^2z(z+c)}{R_2^7} \right]$$

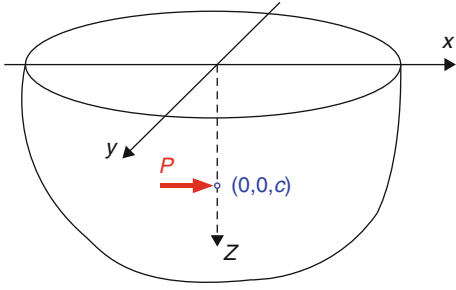
$$\tau_{zz} = \frac{P}{8\pi(1-\nu)} \left[ -\frac{(1-2\nu)(z-c)}{R_1^3} - \frac{3(z-c)^3}{R_1^5} + \frac{(1-2\nu)(z-c)}{R_2^3} - \frac{3(3-4\nu)z(z+c)^2 - 3c(z+c)(5z-c)}{R_2^5} - \frac{30cz(z+c)^3}{R_2^7} \right]$$

$$\tau_{yz} = \frac{Py}{8\pi(1-\nu)} \left[ -\frac{(1-2\nu)}{R_1^3} - \frac{3(z-c)^2}{R_1^5} + \frac{(1-2\nu)}{R_2^3} - \frac{3(3-4\nu)z(z+c) - 3c(3z+c)}{R_2^5} \right]$$

$$\tau_{xz} = \frac{Px}{8\pi(1-\nu)} \left[ -\frac{(1-2\nu)}{R_1^3} - \frac{3(z-c)^2}{R_1^5} + \frac{(1-2\nu)}{R_2^3} - \frac{3(3-4\nu)z(z+c) - 3c(3z+c)}{R_2^5} \right]$$

$$\tau_{xy} = \frac{Pxy}{8\pi(1-\nu)} \left[ -\frac{3(z-c)}{R_1^5} - \frac{3(3-4\nu)(z-c)}{R_2^5} + \frac{4(1-\nu)(1-2\nu)}{R_2^2(R_2+z+c)} \left( \frac{1}{R} + \frac{1}{R_2+z+c} \right) - \frac{30cz(z+c)}{R_2^7} \right]$$

## Horizontal Load



$$u_x = \frac{P}{16\pi\mu(1-\nu)} \left[ \frac{3-4\nu}{R_1} + \frac{1}{R_2} + \frac{x^2}{R_1^3} + (3-4\nu) \frac{x^2}{R_2^5} + \frac{2cz}{R_2^3} \left( 1 - \frac{3x^2}{R_2^2} \right) + \frac{4(1-\nu)(1-2\nu)}{R_2 + z + c} \times \left( 1 - \frac{x^2}{R_2(R_2 + z + c)} \right) \right]$$

$$u_y = \frac{Pxy}{16\pi\mu(1-\nu)} \left[ \frac{1}{R_1^3} + \frac{3-4\nu}{R_2^3} - \frac{6cz}{R_2^5} - \frac{4(1-\nu)(1-2\nu)}{R_2(R_2 + z + c)^2} \right]$$

$$u_z = \frac{Px}{16\pi\mu(1-\nu)} \left[ \frac{z-c}{R_1^3} + \frac{(3-4\nu)(z-c)}{R_2^3} - \frac{6cz(z+c)}{R_2^5} + \frac{4(1-\nu)(1-2\nu)}{R_2(R_2 + z + c)} \right]$$

$$\tau_{xx} = \frac{Px}{8\pi(1-\nu)} \left[ -\frac{1-2\nu}{R_1^3} + \frac{(1-2\nu)(5-4\nu)}{R_2^3} - \frac{3(3-4\nu)x^2}{R_2^5} - \frac{4(1-\nu)(1-2\nu)}{R_2(R_2 + z + c)^2} \times \left( 3 - \frac{x^2(3R_2 + z + c)}{R_2^2(R_2 + z + c)} \right) + \frac{6c}{R_2^5} \left( 3c - (3-2\nu)(z+c) + \frac{5x^2y}{R_2^2} \right) \right]$$

$$\tau_{yy} = \frac{Px}{8\pi(1-\nu)} \left[ \frac{1-2\nu}{R_1^3} + \frac{(1-2\nu)(3-4\nu)}{R_2^3} - \frac{3(3-4\nu)y^2}{R_2^5} - \frac{4(1-\nu)(1-2\nu)}{R_2(R_2 + z + c)^2} \times \left( 1 - \frac{y^2(3R_2 + z + c)}{R_2^2(R_2 + z + c)} \right) + \frac{6c}{R_2^5} \left( c - (1-2\nu)(z+c) + \frac{5y^2z}{R_2^2} \right) \right]$$

$$\tau_{zz} = \frac{Px}{8\pi(1-\nu)} \left[ \frac{1-2\nu}{R_1^3} - \frac{1-2\nu}{R_2^3} - \frac{3(z+c)^2}{R_1^5} - \frac{3(3-4\nu)(z+c)^2}{R_2^5} + \frac{6c}{R_2^5} \times \left( c + (1-2\nu)(z+c) + \frac{5z(z+c)^2}{R_2^2} \right) \right]$$

$$\tau_{zy} = \frac{Pxy}{8\pi(1-\nu)} \left[ -\frac{3(z-c)}{R_1^3} - \frac{3(3-4\nu)(z+c)}{R_2^3} + \frac{6c}{R_2^5} \left( 1-2\nu + \frac{5z(z+c)}{R_2^2} \right) \right]$$

$$\tau_{zx} = \frac{P}{8\pi(1-\nu)} \left[ -\frac{(1-2\nu)(z-c)}{R_1^3} + \frac{(1-2\nu)(z-c)}{R_2^3} - \frac{3x^2(z-c)}{R_1^5} - \frac{3(3-4\nu)x^2(z+c)}{R_2^5} - \frac{6c}{R_2^5} \times \left( z(z+c) - (1-2\nu)x^2 - \frac{5x^2z(z+c)}{R_2^2} \right) \right]$$

$$\tau_{xy} = \frac{Py}{8\pi(1-\nu)} \left[ -\frac{1-2\nu}{R_1^3} + \frac{1-2\nu}{R_2^3} - \frac{3x^2}{R_1^5} - \frac{3(3-4\nu)x^2}{R_2^5} - \frac{4(1-\nu)(1-2\nu)}{R_2(R_2 + z + c)^2} \left( 1 - \frac{x^2(3R_2 + z + c)}{R_2^2(R_2 + z + c)} \right) - \frac{6cz}{R_2^5} \left( 1 - \frac{5x^2}{R_2^2} \right) \right]$$

Other solutions are available for loads on the surface of semi-infinite bodies that are two-dimensional, and that are transversely isotropic and anisotropic, but they will not be reproduced here. Details of Boussinesq-type two-dimensional solutions are given in Johnson (1985).

## Key Applications

A useful application of Green's functions is in computational contact mechanics, as a building block for a semi-analytical solution to a more complicated problem. Although other Green's functions are available for different surface geometries, the above results would be suitable to many contact applications for a surface, when reduced to the Boussinesq or Cerruti solutions (see, e.g., Johnson 1985; Rongved 1955). Johnson also provides the equivalent solutions for two-dimensional Green's functions (plane contact problems).



## Cross-References

- [Contact of Rough Surfaces](#)
- [Crack Growth in Brittle and Ductile Solids](#)
- [Crystal Plasticity with Multiple Slip](#)
- [FFT-Based Methods for Contact Mechanics](#)
- [Fractal Geometry](#)

## References

- J.R. Barber, *Elasticity* (Springer, Dordrecht/New York, 2010)
- J. Boussinesq, *Application des potentiels à l'étude de l'équilibre et du mouvement des solides élastiques* (Cauthier-Villars, Paris, 1885)
- A.E. Green, W. Zerna, *Theoretical Elasticity* (Clarendon, Oxford, 1968)
- D.A. Hills, P.A. Kelly, D.N. Dai, A.M. Korsunsky, *Solution of Crack Problems* (Kluwer, Dordrecht/Boston, 1996)
- K.L. Johnson, *Contact Mechanics* (Cambridge University Press, Cambridge/New York, 1985)
- R.D. Mindlin, Force at a point in the interior of a semi-infinite solid. *Physics* 7, 195–202 (1936)
- T. Mura, *Micromechanics of Defects in Solids* (Kluwer, Dordrecht/Boston/London, 1993)
- L. Rongved, Force at a point in the interior of one of the two joined semi-infinite solids, in *Proceedings of 2nd Midwest Conference Solid Mechanics*, Lafayette (1955)

## Grid-Convergence of EHL Solution

- [Mesh Density Effect on EHL Solution](#)

## Griffith Theory of Fracture

ALAN T. ZEHNDER

Field of Theoretical and Applied Mechanics,  
Cornell University, Ithaca, NY, USA

### Synonyms

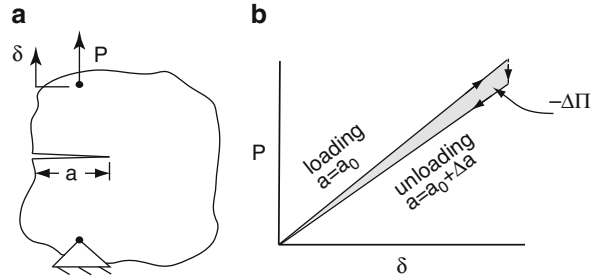
[Brittle fracture theory](#); [Linear elastic fracture](#)

### Definition

The Griffith theory states that a crack will propagate when the reduction in potential energy that occurs due to crack growth is greater than or equal to the increase in surface energy due to the creation of new free surfaces. This theory is applicable to elastic materials that fracture in a brittle fashion.

### Scientific Fundamentals

Through a series of experiments, stress analyses, and synthesis of prior work, in his remarkable paper Griffith



**Griffith Theory of Fracture, Fig. 1** (a) A cracked, linear elastic sheet under single point loading. (b) Load–displacement plot assuming that the displacement is held constant as the crack grows by increment  $\Delta a$ . The shaded region indicates the reduction in elastic energy of the body due to crack growth

(1921) developed the fundamental concept that underlies the modern theory of linear elastic fracture mechanics. His theory is based on balancing the reduction of potential energy that occurs during fracture with the increase in surface energy due to the creation of new free surfaces when a crack grows.

The energy release rate,  $G$ , is defined as the energy that flows to the crack tip per unit of new crack surface created. An energy balance shows that

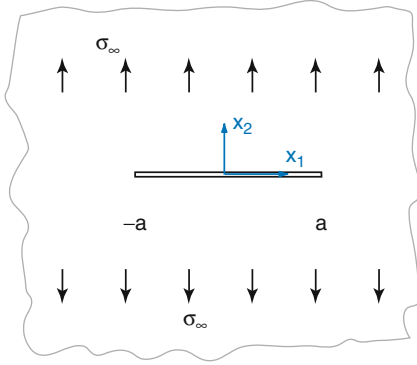
$$G = -\frac{\partial \Pi}{\partial s}, \quad (1)$$

where  $s$  is the fracture surface and  $\Pi$  is the potential energy given by

$$\Pi = \int_V W dV - \int_{S_t} \mathbf{t} \cdot \mathbf{u} dS - \int_V \mathbf{b} \cdot \mathbf{u} dV, \quad (2)$$

where  $W$  is the strain energy density,  $V$  is the region occupied by a body,  $S_t$  is the surface of the body upon which tractions,  $\mathbf{t}$ , are prescribed,  $\mathbf{u}$  is the displacement field, and  $\mathbf{b}$  is the body force (Rice 1968).

As a simple illustration of the energy release rate, consider a cracked, linearly elastic sheet of thickness  $B$ , loaded by a prescribed displacement at a single point as shown in Fig. 1a. Assume that no body forces act on this sheet. The applied displacement is  $\delta$  with reaction force  $P$ . If the body with initial crack length  $a_0$  is loaded, the  $P - \delta$  curve will be linear as shown in Fig. 1b. The total strain energy of the body,  $\int_V W dV$ , is the area under the  $P - \delta$  curve. Suppose that at some point the crack grows to a new length  $a_0 + \Delta a$  and that during this crack extension process the displacement is held constant. As the crack grows the load will drop. If the body is unloaded after crack extension,  $-\Delta\Pi$ , the shaded region in Fig. 1b



**Griffith Theory of Fracture, Fig. 2** Through crack of length  $2a$  in a plate under tension. A uniform stress field  $\sigma_{22} = \sigma_{\infty}$  is assumed to exist away from the crack

represents the energy dissipated by crack growth. The energy “released” by the body per unit new fracture surface is

$$G = - \lim_{\Delta a \rightarrow 0} \frac{1}{B} \frac{\Delta \Pi}{\Delta a}.$$

The Griffith theory postulates that a crack will grow when the energy released by the body per unit fracture area is greater than or equal to the increase in surface energy,

$$G \geq 2\gamma, \quad (3)$$

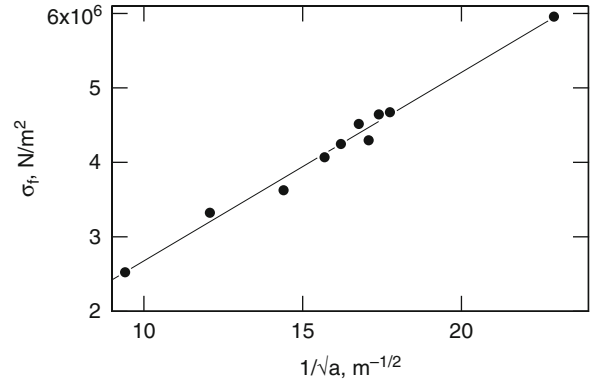
where  $\gamma$  is the surface energy. The canonical problem of the Griffith theory is the through crack in a plate under tension, as shown in Fig. 2. Assuming plane stress,  $G$  for this problem is (Rice 1968)

$$G = \frac{\sigma_{\infty}^2 \pi a}{E}, \quad (4)$$

where  $\sigma_{\infty}$  is the tensile stress,  $E$  is the Young’s modulus, and  $a$  is the half crack length. Setting  $G = 2\gamma$  and solving for  $\sigma_f$ , the predicted stress at the onset of fracture,

$$\sigma_f = \sqrt{\frac{2\gamma E}{\pi a}}. \quad (5)$$

To test the theory, Griffith had glass taken from test tubes blown into thin-walled spheres and cylinders. A glass cutter was used to introduce through cracks in the test samples, which were then annealed to eliminate any residual stresses due to cutting. The samples were then pressurized and the pressure at the point of unstable fracture recorded. The stress  $\sigma_f$  at fracture was then calculated for each test. The results, shown in Fig. 3, demonstrate that the failure stress is linear with  $1/\sqrt{a}$  as predicted by the theory. However, the measured values of



**Griffith Theory of Fracture, Fig. 3** Fracture stress vs.  $1/\sqrt{a}$  from Griffith’s experiments on cracked glass tubes and spheres.  $a$  is the half crack length, and  $\sigma_f$  is the measured failure stress. Data are shown as points. The solid line is a linear fit to the data. Note that the fracture strength will not increase indefinitely as  $a$  get smaller. The strength of a “perfect” structure is limited by the intrinsic bond strength of the material. In the case of glass the theoretical maximum strength is on the order of 1 GPa, thus Griffith’s measured fracture strengths are about an order of  $100\times$  lower than the theoretical strength. The sensitivity of glass and other brittle materials to small flaws is what ultimately limits their practical strengths. Thin fibers, which due to their very small dimensions cannot contain large cracks, will generally exhibit greatly improved tensile strengths relative to bulk glass

$\sigma_f$  are larger than predicted by Griffith. Griffith estimated the surface energy as  $\gamma = 0.54 \text{ J/m}^2$ . The data in Fig. 3 are best fit with  $\gamma = 1.62 \text{ J/m}^2$ . Where does the difference come from? Griffith’s  $\gamma$  is extrapolated down to room temperature based on  $\gamma$  values measured for molten glass at 745–1,110°C. The extrapolation may well be incorrect.

However, surface energy does not account for all the energy of fracture. Inelastic deformations such as plastic flow, void growth, microcracking, and crazing dissipate energy in the “process zone” surrounding the crack tip. The generalization of the Griffith theory is that a crack grows when

$$G \geq 2\gamma + U_f,$$

where  $U_f$  is the inelastic energy dissipation per unit of crack growth. In a great many materials,  $U_f \gg 2\gamma$  and the fracture criterion is stated simply as

$$G \geq G_C, \quad (6)$$

where  $G_C$  is a critical value of the crack tip energy release rate. The value of  $G_C$  can be measured using a number of

standard tests in which a cracked sample is loaded to failure and the fracture energy at failure computed (see, e.g., ASTM 2005 or Janssen et al. 2004). In brittle materials  $G_C$  is generally considered to be a material property, although in practice it will depend on temperature, environment, and rate of loading and may not be constant as the crack grows. Note that the Griffith theory can be further generalized to consider any cracked body. It need not be a finite crack in a plate as shown in Fig. 2.

## Key Applications

### Prediction of Fracture

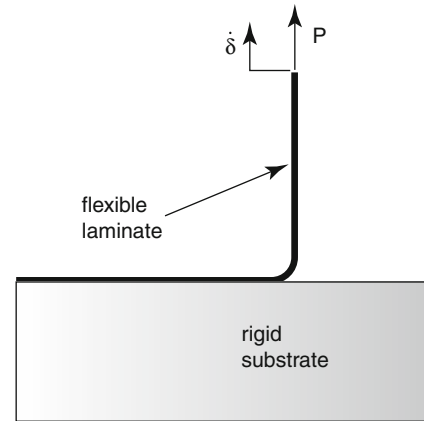
As has already been stated, the Griffith theory is useful for predicting the initiation of crack growth from a sharp crack in a plate under tensile loading. If the crack is loaded with tensile and shear forces (mixed-mode loading) so that far from the crack  $\sigma_{22} = \sigma_\infty$  and  $\sigma_{12} = \tau_\infty$  and if the crack were to grow straight ahead, the energy release rate would be

$$G = \frac{\sigma_\infty^2 \pi a}{E} + \frac{\tau_\infty^2 \pi a}{E}. \quad (7)$$

However, under mixed-mode loading the crack will generally not grow straight ahead. Extending Griffith's ideas, a criterion for predicting the angle of crack growth can be developed by postulating that the crack will grow at an angle that maximizes the energy release rate, hence minimizing the total system potential energy. Wu (1978) considered a straight crack as shown in Fig. 2 but with a kink at an angle  $\theta$  from the  $x_1$  axis. Calculating the energy of the kinked crack in the limit as the kink length goes to zero, Wu calculates  $G(\theta)$ , the energy release rate that would occur if the crack kinked along angle  $\theta$ . Taking the maximum value of  $G(\theta)$  for different mixes of  $\sigma_\infty$  and  $\tau_\infty$  allows one to predict the angle of crack growth for mixed-mode loading. The results show that the crack will evolve to a direction perpendicular to the maximum principal stress of the plate. In an isotropic material the theory of maximum energy release rate can be extended to predict the growth of cracks in three dimensions. In this case, a crack could propagate for different distances and in different directions at various points along the crack front. Out of all possible fracture surfaces, the one that maximizes the energy release rate is the new fracture surface (Hwang 1999).

### Adhesion Energy

Although Griffith theory is formulated to predict fracture of a material, it has been extended to a great many problems. As just one example of how Griffith theory has been applied to a practical problems, consider adhesion. Adhesive bonds



**Griffith Theory of Fracture, Fig. 4** Characterization of adhesion through measurement of the energy required to debond one adherent from another is an example of the application of Griffith theory to problems never envisioned by Griffith. Shown is the 90° peel test in which a laminate of width  $B$  is peeled from a substrate by force  $P$ . The peel test can be performed for other angles as well

can be found in many situations, such as films on a substrate, a bonded lap joint, or an electronic package consisting of several bonded layers of dissimilar materials. The quality of an adhesive bond can be characterized in terms of the strength of the bond or in terms of the energy per unit area required to disbond the adhered materials.

As an example, consider the adhesion of a flexible laminate on a rigid substrate. In the 90° peel test shown in Fig. 4, a laminate of width  $B$  is peeled from the substrate by force  $P$ . The rate of work input to the system is the force,  $P$ , times the load-point displacement rate,  $\dot{\delta}$ . Assuming that the laminate has negligible stretch, the rate of increase of fracture surface area is  $B\dot{\delta}$ . Balancing the rate of work done by the rate of increase of fracture energy,  $P\dot{\delta} = B\dot{\delta}G$ , hence the fracture energy is given by  $G = P/B$  for this simple case. Moore and Williams (2001) provide further details and generalizations.

### Limitations

As mentioned above, if the Griffith theory is generalized to include not just surface energy but the process zone energy it can be extended to predict fracture in a great many materials, provided that the extent of inelastic deformation is not large relative to the dimensions of the body. In cases of unconfined plastic deformation the use of  $G$  as a fracture criterion may not be accurate and other approaches must be used. Even in cases where  $G$  can be used to predict the onset of crack growth, as a crack grows

and leaves a wake of plastic deformation, the value of  $G$  needed to continue to grow the crack may change rapidly and one may not be able to predict continued crack growth based on a critical  $G$  value.

## Cross-References

- [Crack Growth in Brittle and Ductile Solids](#)
- [Crack Growth in Noncrystalline Solids](#)
- [Crack Initiation in Brittle Solids](#)
- [Modes of Fracture](#)
- [Stress Intensity Factors](#)

## References

- ASTM, *ASTM E 1820: Standard Test Method for Measurement of Fracture Toughness* (ASTM International, West Conshohocken, PA, 2005)
- A.A. Griffith, The phenomena of rupture and flow in solids. *Philos. Trans. R. Soc. Lond. A* **221**, 163–198 (1921)
- C.W. Hwang, Virtual crack extension method for calculating rates of energy release rate and numerical simulation of crack growth in two and three dimensions. Ph.D. thesis, Cornell University 1999
- M. Janssen, J. Zuidema, R. Wanhill, *Fracture Mechanics* (Spon Press, London, 2004)
- D.R. Moore, J.G. Williams, Peel testing of flexible laminates, in *Fracture Mechanics Testing Methods for Polymers, Adhesives and Composites* (Elsevier, Amsterdam, 2001), pp. 203–224
- J.R. Rice, Chap 3. Fracture, in *Mathematical Analysis in the Mechanics of Fracture*, ed. by H. Liebowitz, vol. II (Academic Press, New York, 1968), pp. 191–311
- C.H. Wu, Maximum energy release rate criterion applied to a tension-compression specimen with crack. *J. Elasticity* **8**, 235–257 (1978)

## Grinding for Gear Manufacturing

- [Gear Manufacturing Machines](#)

## Grinding Lubricants

- [Metalworking Lubricants](#)

## Grinding Processes

STEPHEN MALKIN

Department of Mechanical & Industrial Engineering,  
University of Massachusetts, Amherst, MA, USA

## Synonyms

[Abrasive machining processes](#)

## Definition

Grinding is the common collective name for machining processes that utilize hard, abrasive particles as the cutting medium. The grinding process of shaping materials is probably the oldest in existence, dating from the time prehistoric humans found that they could sharpen their tools by rubbing them against gritty rocks.

## Scientific Fundamentals

### Types of Grinding Operations

There are numerous types of grinding operations, which vary according to the shape of the wheel and the kinematic motions of the workpiece and wheelhead. Some of the more common ones for machining of flat and cylindrical surfaces are illustrated in [Fig. 1](#). More complex machines are used to generate other shapes. Any of these processes may be applied to fine finishing, to large-scale stock removal, or to a host of tasks between these extremes.

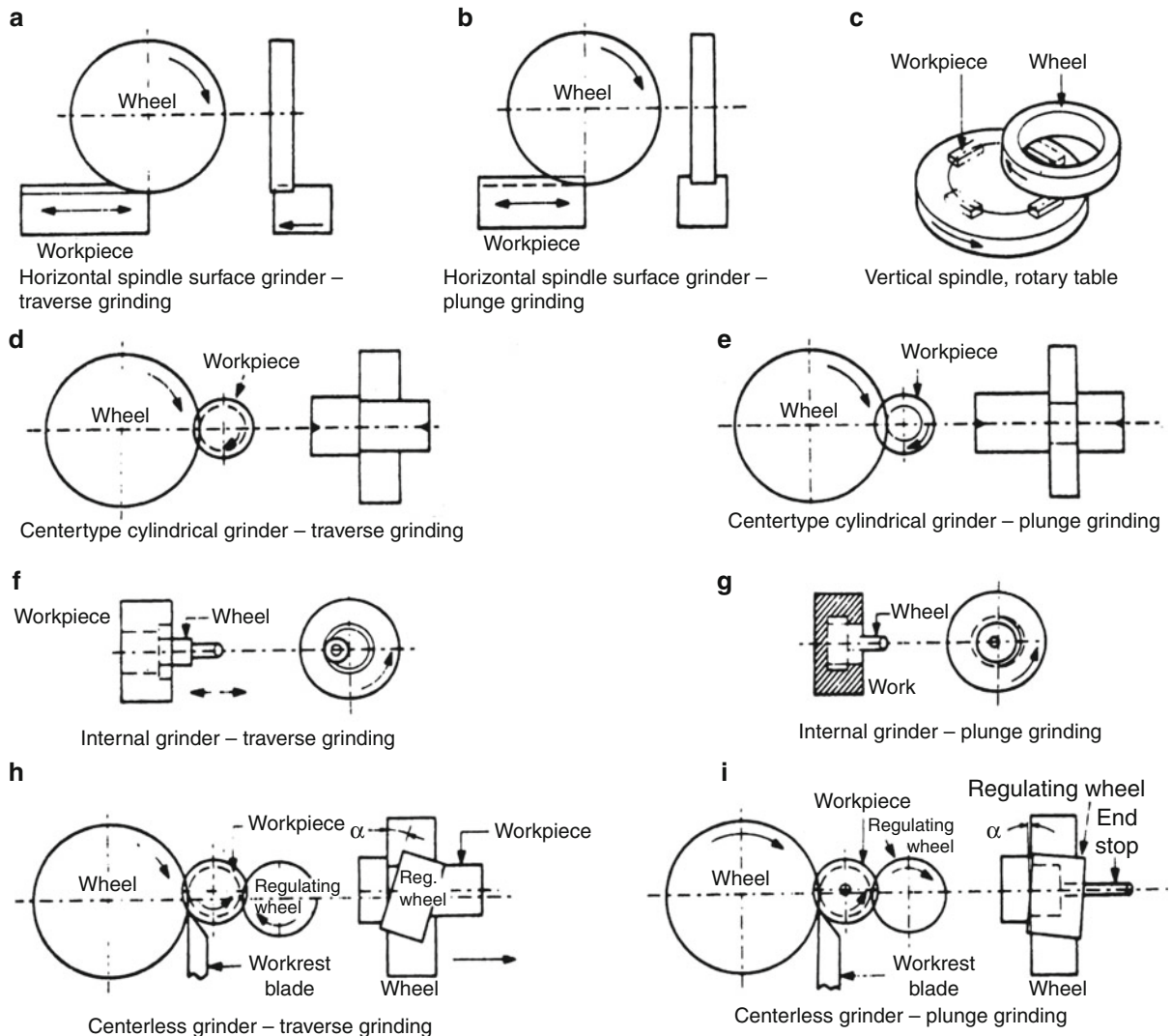
### Grinding Wheels

Prehistoric humans' abrasive tool was natural sandstone, which contains grains of sand in a silicate bond matrix. Modern grinding wheels are fabricated by cementing together tiny abrasive grains or grits, which do the cutting, with a softer bonding material. Depending on the particular type of bond, the space between the abrasive particles may only be partially filled, leaving gaps and porosity, or completely filled with binder. Aside from abrasive and bond material, fillers and grinding-aid materials may also be added.

Grinding wheels are made from many types of grit in a wide range of sizes, in conjunction with many bond materials and compositions. "Conventional" wheels in common use contain either aluminum oxide or silicon carbide abrasive with vitrified or resinoid bonds. "Superabrasive" wheels with diamond and cubic boron nitride (CBN) abrasives are produced with vitrified, resin, and metal bonds. Conventional abrasive wheels usually comprise the entire bonded abrasive structure throughout, but the abrasive-composite on ultrahard, superabrasive wheels is limited to a thin rim or layer on a solid hub in order to reduce the amount of costly diamond and CBN. The different types of grinding wheels, together with the requirements for wheel shapes and sizes to fit all the diverse grinding machines and jobs to be done, lead to an almost endless diversity of grinding wheels.

### Wheel Preparation

With the exception of some heavy-duty and cut-off operations, the grinding wheel topography and the



**Grinding Processes, Fig. 1** Illustration of grinding operations for machining flat and cylindrical surfaces (Kalpakjian 1984)

macroscopic wheel shape are generated by preparing the wheel prior to grinding and periodically during the course of grinding. Wheel preparation generally includes truing and dressing. Truing usually refers to removal of material from the cutting surface of a grinding wheel so that the spinning wheel runs true with minimum run-out from its macroscopic shape; truing may also include profiling of the wheel to a particular shape. Dressing is the process of conditioning the wheel surface so as to achieve a certain grinding behavior. With conventional abrasive wheels, both truing and dressing are usually done by the same process, and the combination is commonly called dressing.

Grinding wheels containing conventional ceramic abrasives are usually prepared by feeding a diamond dressing tool across the rotating wheel surface. During each pass of the dressing tool across the wheel face, a shallow layer is removed from the wheel periphery. This type of dressing motion is analogous to turning on a lathe. Another dressing method that is especially used for generating profiles is rotary diamond dressing, which utilizes an axisymmetric dressing tool (roll) consisting of diamond particles impregnated in a metal matrix or held by an electroplated metal coating on its outer surface.

Virtually all types of diamond and CBN superabrasive wheels undergo pre-grinding preparation by truing, which

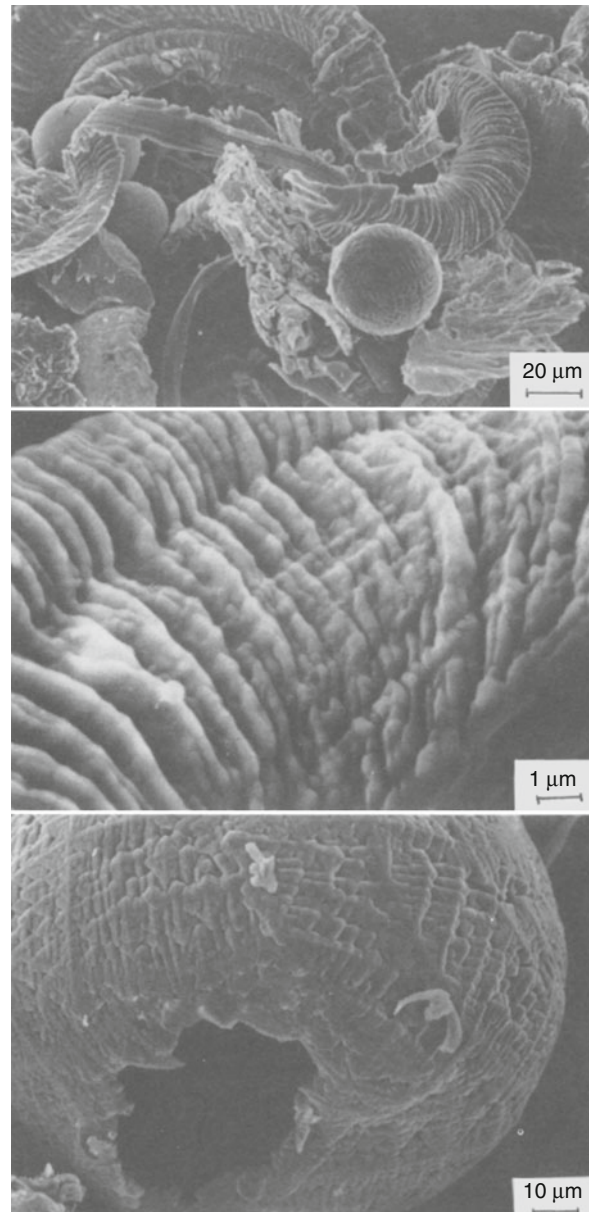


may be followed by dressing. An exception arises in the case of single-layer electroplated wheels, which may only require occasional “cleaning” or “touching up” with an abrasive stick. There are various types of dressing processes that utilize vitrified silicon carbide wheels, stationary multipoint diamond tools, or rotary diamond tools. CBN is somewhat softer than diamond, so diamond truing tools tend to wear much less than diamond wheels. After truing, dressing of diamond wheels and resin and metal bonded CBN wheels is usually accomplished by infeeding a fine-grained vitrified abrasive stick into the wheel surface.

### Grinding Mechanisms

Material removal during grinding occurs as abrasive grains interact with the workpiece. The penetration depths of the cutting points on the rotating wheel into the material being ground depend upon the topography of the wheel surface and the geometry and kinematic motions of the wheel and workpiece. For grinding of metals, material removal occurs mainly by a shearing process of chip formation, similar to that found with other machining methods such as turning or milling. In Fig. 2, a scanning electron micrograph of the debris (swarf) recovered after grinding a plain carbon steel reveals mostly curled chips, very much like those produced by turning or milling, but somewhat irregular in size and shape owing to the variability in cutting-point geometry and penetration depths. These chips have a very fine lamella structure, as seen in the middle SEM photo. Two other types of particles found in the grinding swarf are short segmental blocky chips and spheres. Blocky-chip formation occurs by an extrusion-like bulging process with extreme negative rake angles. An isolated spherical grinding particle shown at the bottom in Fig. 2 has a hollow shape and an extremely fine dendritic microstructure, indicating that it was once molten and solidified rapidly. Melting does not usually occur during grinding, but only afterwards by exothermic reaction of small hot chips with oxygen in the atmosphere. The round and hollow shape is attributed to surface tension effects acting on the molten curled chip. Oxidation of grinding chips as they are emitted from the wheel is responsible for the spark stream, which is actually glowing chips.

A fundamental parameter derived from the power and machining conditions is the specific energy, which is defined as the energy expended per unit volume of material removal. As compared with larger-scale machining processes, grinding requires a specific energy that is much bigger, typically by one or two orders of magnitude. This is mainly due to plowing and sliding processes, which



**Grinding Processes, Fig. 2** SEM micrographs of grinding debris from a medium carbon steel workpiece

occur in addition to chip formation in grinding. Plowing energy is expended by deformation of workpiece material without removal, including side flow of material from the cutting path into ridges and plastic deformation of the material passing under the cutting edge. Sliding involves rubbing of dulled flattened wear flats areas on the abrasive grains against the workpiece surface without removing any material. Such “wear flats” are generated by dressing



prior to grinding, and become glazed and further enlarged during grinding through attritious wear and adhesion of metal particles from the workpiece. The tendency for wheel dulling may be offset, to a greater or lesser degree, by “self-sharpening” due to bulk wear of the wheel by grit fracture or grit dislodgement from the binder.

Only the specific chip-formation energy is actually expended by material removal and, as such, is the minimum grinding energy required. However, the chip formation energy in grinding is still much bigger than in larger-scale metal-cutting operations. This is a consequence of the extremely large negative rake angles at the cutting points on abrasive grains, which are typically  $-60^\circ$  or even more negative. For such large negative rake angles, classic chip-formation theory predicts extremely large shear strains. Such large strains are obtained at extremely high cutting velocities and strain rates so that the plastic deformation is nearly adiabatic. For adiabatic deformation to such large strains, the plastic deformation energy in grinding is comparable to the melting energy of the workpiece material.

### Thermal Aspects

The grinding process requires an extremely high energy expenditure per unit volume of material removed. Virtually all of this energy is converted to heat, which is concentrated within the grinding zone. The high temperatures produced can cause various types of thermal damage to the workpiece, such as burning, phase transformations, softening (tempering) of the surface layer with possible rehardening, unfavorable residual tensile stresses, cracks, and reduced fatigue strength. Furthermore, thermal expansion of the workpiece during grinding contributes to inaccuracies and distortions in the final product. The production rates that can be achieved by grinding are often limited by grinding temperatures and their deleterious influence on workpiece quality.

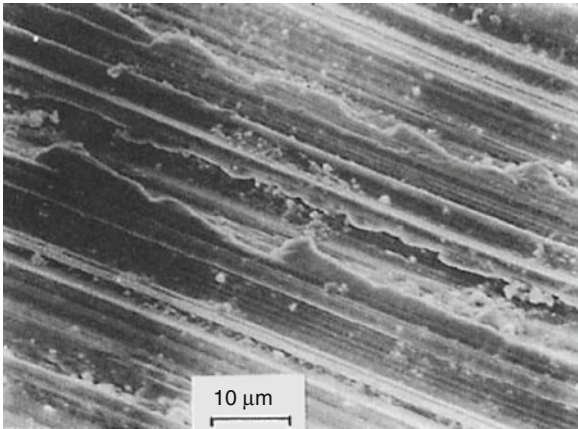
During grinding, peak “flash” temperatures are generated that approach the melting point of the material being ground. However, these extreme temperatures are of extremely short duration and highly localized on the shear planes of microscopic grinding chips. Just beneath the surface, the workpiece “feels” nearly continuous heating owing to the multiplicity of interactions with the abrasive grits passing quickly through the grinding zone. The temperature associated with “continuous” heating over the grinding zone, rather than the peak “flash” temperature, is found to be responsible for most thermal damage. Also of interest is the temperature in the bulk of the workpiece, which causes thermal expansion leading to distortions and dimensional inaccuracies.

Thermal analyses of grinding processes are usually based upon the application of moving heat source theory to the workpiece being ground. For this purpose, the grinding zone is usually modeled as a planar heat source that moves along the surface of the workpiece. A critical factor for calculating grinding temperatures is the energy partition, which is the fraction of the grinding energy transported as heat to the workpiece at the grinding zone. For shallow cut grinding with conventional abrasive wheels, the energy partition is typically 60–85%. Much smaller energy partitions of only 3–7% are obtained for creep-feed grinding with slow work speeds and large depths of cut. Such low energy partitions are attributed mainly to cooling by the fluid at the grinding zone, which is further enhanced by the porous wheel structure. Heat conduction to the grains can also reduce the energy partition especially with CBN abrasives that have high thermal conductivity. For high-efficiency deep grinding (HEDG) using CBN wheels with large depths of cut and fast work speeds, preheated material ahead of the grinding zone is removed together with the chips, thereby lowering the temperature on the finished surface. Analytical models have been developed that take all of these effects into account and provide the technological basis for prediction and control of thermal damage.

### Surface Roughness and Tolerance

Grinding processes are often selected for final finishing of components because of their ability to satisfy stringent requirements of surface roughness and tolerance. Surface roughness and tolerance are closely interrelated, as it is generally necessary to specify a smoother finish in order to maintain a finer tolerance in production. Typical arithmetic average surface roughnesses for production grinding operations range from about 0.15 to 1.5  $\mu\text{m}$ , although finishes outside this range are not uncommon. Corresponding dimensional tolerances specified in production are normally 10–50 times the arithmetic average roughness.

The fine-scale morphology of the surfaces generated by grinding consists mostly of overlapping scratches produced by the interaction of abrasive cutting points with the workpiece. An example of a typical ground surface is shown in Fig. 3. For this example of straight plunge grinding, as with other types of grinding, the grit motion relative to the workpiece is readily identified from the directionality of the scratches and grooves. Sideways displacement of material from some scratches by plowing is also evident. The ground surface morphology is further complicated by numerous other phenomena. Back transfer of workpiece metal often occurs, especially with



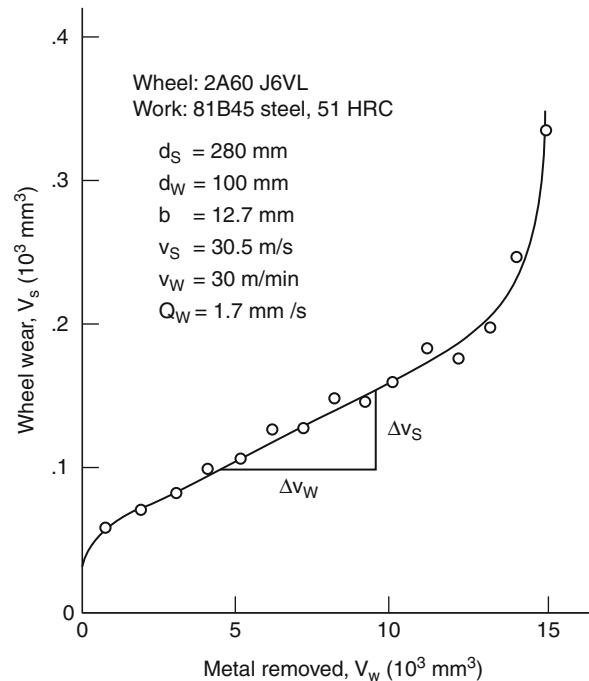
**Grinding Processes, Fig. 3** SEM micrograph of a medium carbon steel workpiece surface after straight surface grinding with a vitrified aluminum oxide wheel and soluble oil

adhesive metals, whereby metal particles adhering to the abrasive grits are re-deposited on the workpiece. Interruption of the cutting action by fracture of the abrasive grit may leave a crater on the workpiece, possibly with an abrasive fragment embedded in the surface.

As with other machining processes, the surface roughness for grinding can be theoretically predicted by modeling how the abrasive cutting points on the rotating wheel surface kinematically interact with the workpiece. For this purpose, the surface topography is assumed to be generated by clean cutting, whereby the cutting edges remove all material that they encounter in their paths, leaving behind the resulting cutting grooves. However, the theoretical roughness is usually found to be much less than the actual roughness due to such factors as sideflow plowing and vibrations. Theoretical analyses of surface roughness provide physical insight into how ground surfaces are generated and what may be the controlling factors. However, it is generally necessary to rely upon empirical relationships to quantitatively predict how the grinding and dressing parameters affect the actual surface roughness.

### Wheel Wear

The wear of a grinding wheel is usually expressed as a volumetric loss of material. An illustration of typical wheel-wear behavior is shown in Fig. 4 as a plot of volumetric wheel wear versus accumulated metal removed  $V_w$ . Since the removal rate is essentially constant, the accumulated metal removed on the horizontal axis is proportional to time. The wear behavior seen here is similar to that observed with other wear processes. High initial wear is

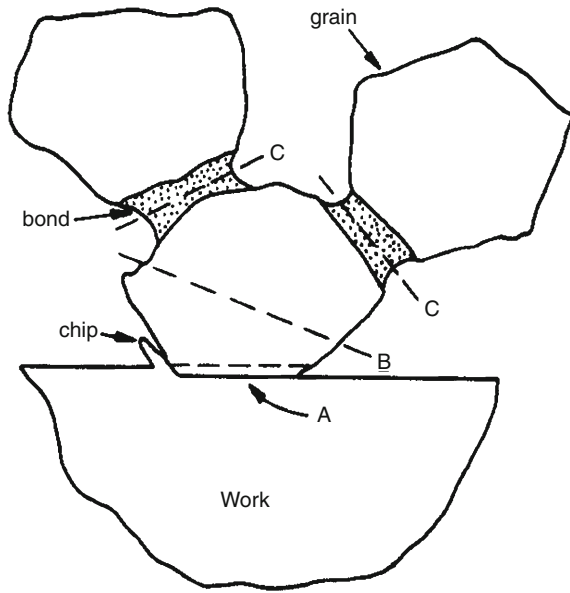


**Grinding Processes, Fig. 4** Volumetric wheel wear versus accumulated metal removed for a cylindrical grinding operation

followed by a steady-state regime with a nearly constant wear rate. A third regime of accelerating wear, as seen in Fig. 4, usually indicates a “catastrophic” situation and the need to redress the wheel. Accelerating wear, if and when it occurs, may be associated with workpiece burn or chatter.

A performance index commonly used to characterize wheel-wear resistance is the “grinding ratio,” also referred to as the G-ratio or  $G$ , which is the volume of material removed per unit volume of wheel wear. This parameter may be computed for the entire test or only for the steady-state wear regime. G-ratios cover an extremely wide range of values. On vanadium-rich high-speed steels, G-ratios less than unity may be obtained, in which case the work appears to be abrading the wheel rather than vice versa. At the other extreme, G-ratios of above 60,000 may be obtained using CBN wheels with a straight oil as the grinding fluid.

It is generally recognized that there are three main mechanisms of wheel wear as illustrated in Fig. 5: attritious wear, grain fracture, and bond fracture. Attritious wear involves dulling of abrasive grains and the growth of wear flats by rubbing against the workpiece. Grain fracture refers to removal of abrasive fragments by



**Grinding Processes, Fig. 5** Illustration of wheel-wear mechanisms: A – attritious wear, B – grain fracture, and C – bond fracture

fracture within the grain, and bond fracture to dislodging of the abrasive grain from the binder. Both these types of fracture wear may lead to self-sharpening, which reduces the dulled wear flat area caused by attritious wear.

The suitability of abrasive grain materials for grinding particular workpiece metals depends to a very large degree on their attritious wear resistance. Where possible, the abrasive should be considerably harder than the material being ground, but hardness by itself is often not the prevailing factor. If it were, diamond would be the most wear-resistant abrasive for grinding any metal, and silicon carbide would rank higher than aluminum oxide. In practice, neither diamond nor silicon carbide is the best choice for grinding most ferrous alloys.

Attritious wear and the dulling of abrasive grits are both chemical and mechanical. Chemical effects are likely to be more significant when the abrasive is somewhat harder than the workpiece and any of its included phases. As an abrasive grain interacts with the workpiece at the elevated temperatures reached in the grinding zone, numerous chemical reactions may occur involving the abrasive, workpiece metal, binder, atmosphere, and grinding fluid in various combinations. For example, diamond, despite its extreme hardness, is not suitable for grinding most ferrous alloys. This anomalous behavior can be attributed to excessive attritious wear mainly by the

reversion of diamond to graphite which is catalyzed in the presence of iron. Cubic boron nitride (CBN), although somewhat softer than diamond, is more chemically stable to higher temperatures and wears much less on most ferrous metals. For grinding ferrous metals with aluminum oxide wheels, the important chemical reactions usually involve the oxidation of iron and the reaction of the oxide with the abrasive to form the spinel  $\text{FeAl}_2\text{O}_4$ , which is an intermediate compound between the oxidized workpiece metal and aluminum oxide linked to stronger bonding between iron and aluminum oxide. Yet, despite the apparent role of oxygen in promoting adhesion and attrition, its elimination by grinding in a vacuum or inert atmosphere has a drastic effect on the process as nascent metal workpiece particles and freshly formed uncontaminated surfaces tend to physically adhere to each other and to the wheel surface, thereby loading and clogging the wheel. Surface oxidation and corrosion in a normal grinding atmosphere reduce the adhesion of metal particles to each other and their re-adhesion to the workpiece.

Aside from chemical activity, purely mechanical factors contribute significantly to attritious wear. In grinding some carbon and alloy steels, the G-ratio is usually found to be reduced somewhat when grinding the material in its fully hardened condition as compared with its annealed state, which would suggest a mechanical effect. But hardness, by itself, is not necessarily indicative of grindability and attritious wear for grinding of materials, including those whose hardest phases are softer than the aluminum oxide abrasive. A rather different situation arises when grinding high-speed tool steels. Dispersed carbide phases of tungsten, molybdenum, and vanadium in these materials are hard enough to cut or shatter the aluminum oxide and cause very low G-ratios, which accounts for the particular success of the much harder CBN, compared with aluminum oxide, in grinding high-speed steels. The hardest carbides in high-speed steels are of tungsten, molybdenum, and vanadium.

### Grinding Fluids and Lubrication

Many grinding operations are performed with the aid of a grinding fluid. The grinding fluid is generally considered to have two main roles: cooling and lubrication. Grinding fluids are commonly referred to as coolants, but their role as lubricants is often more important.

Most grinding fluids can be categorized either as straight (or neat) oils or as soluble oils. Straight oils for grinding are mineral-oil-based fluids with additions of fatty materials for lubrication and wettability, and usually sulfur and/or chlorine for added wear reduction. Soluble

oils are water-based fluids containing oil emulsions and numerous other ingredients which may include fatty materials, soaps, sulfur, and chloride for lubrication, surfactants for wetting and detergency and to prevent foaming, rust inhibitors, water conditioners, and germicides. Straight oils are generally found to be better lubricants than soluble oils, as evidenced by higher G-ratios, lower grinding forces, and better surface quality. Their superior performance appears to be related mainly to their ability to reduce attritious wear, which also reduces the forces and power to much lower forces as well.

In view of the importance of chemical reactions on attritious wear, it seems likely that the influence of the grinding fluid as a lubricant may be related to how it affects the grinding chemistry. The formation of lubricating films, either by chemical or by physical action, can reduce workpiece-metal adhesion and also inhibit those chemical reactions that promote attritious wear. For grinding of steels with aluminum oxide abrasives, lubricating films obtained with an active straight oil can inhibit spinel formation and metal adhesion. For grinding of steels with silicon carbide, the straight oil may reduce attritious wear in a similar way by inhibiting the dissociation of silicon carbide and the diffusion of carbon to the workpiece.

Commercial grinding fluids typically contain sulfur and/or chlorine as lubrication additives. Analogous to what has been postulated regarding the role of these elements in boundary lubrication of metals, it is generally believed that sulfur and chlorine react with the metal to form sulfide and chloride lubricating films, although the actual reactions seem to be much more complex. Beneficial lubricating effects of sulfur are also realized when it is incorporated as an addition to the workpiece (e.g., sulfurized free-machining steels).

Cooling is most efficient with water-based fluids and lubrication with straight oils. With the notable exception of creep-feed grinding, cooling by grinding fluids appears to be generally ineffective in lowering the peak temperature within the grinding zone. With improved lubrication and reduced wheel dulling, the grinding forces are reduced, thereby lowering the grinding zone temperature and the tendency for thermal damage. These considerations would seem to weigh heavily in favor of straight oils as opposed to soluble oils. But in actual practice, straight oils are used much less than water-based fluids. One advantage of water-based fluids as coolants is their superior ability to control the bulk temperature of the workpiece, which can reduce part-to-part size variations associated with thermal deformation of the workpiece, but this seems to be a secondary

factor favoring the use of soluble oils. Unless lubrication is critical for form and finish, straight oils tend to be avoided mainly because of pollution and safety considerations. Oil-based fluids create mist and fumes in the atmosphere, and they may also present a fire hazard. Special environmental and safety precautions and equipment are necessary.

For a grinding fluid to be effective as a lubricant and coolant, it must be delivered in sufficient quantity to the grinding zone. There are two main methods of grinding fluid delivery: low pressure (flood) and high pressure by means of nozzles. Most conventional grinders are equipped with simpler low-pressure systems, whereas creep-feed and high-speed grinders are more likely to have high-pressure systems. For creep-feed grinding at high rates of stock removal, large quantities of heat must be removed by pumping large volumes of fluid through the grinding zone. With high-speed grinding, the fluid velocity with a low-pressure system may be insufficient to penetrate the boundary layer of air surrounding the wheel, thereby preventing the fluid from reaching the grinding zone. Aside from using high pressures, one simple solution to this problem is to break up the air film by positioning a scraper plate close to the wheel surface at a location just ahead of where the fluid hits the wheel. With high-pressure systems, multiple nozzles are often used to enhance fluid flow to the grinding zone, and this may have the added benefit of reducing the sensitivity of the system to nozzle location and orientation. In addition to their roles as coolants and lubricants, grinding fluids applied at high pressure also mechanically clean the wheel surface by removing adhered metal. Separate nozzles for wheel cleaning may be directed towards areas on the wheel surface away from the grinding zone.

Grinding fluids present environmental and safety hazards, so there is considerable interest in reducing the amount of fluid applied. One approach has been to apply minimum quantity lubrication (MQL), whereby a miniscule amount of a non-hazardous ester oil is applied at a controlled rate using a precision dispenser. MQL has been found to provide comparable or better lubricating performance than soluble oil. However, one possible drawback with MQL is insufficient bulk cooling of the workpiece, which may lead to dimensional errors with long grinding cycles.

## Key Applications

Grinding is a major manufacturing process that accounts for about 20–25% of the total expenditures on machining operations in industrialized countries. Society, as we know it, would be quite impossible without grinding. Almost

everything that we use has either been machined by grinding at some stage of its production, or has been processed by machines that owe their precision to abrasive operations. Grinding is an essential process in the manufacture of virtually all types of mechanical equipment and cutting tools. For example, the production of rolling element bearings, automotive components, power generation equipment, and computer and electronic components is critically dependent on grinding processes.

Although grinding is traditionally regarded as a final machining process in the production of components requiring smooth surfaces and fine tolerances, much more abrasive is actually consumed by heavy-duty grinding operations where the objective is to quickly remove material with little concern for surface quality. The grinding process is also essential for machining of materials that, because of their extreme hardness or brittleness, cannot be efficiently shaped by other methods. Machining of non-metallic brittle materials, including ceramics, cemented carbides, and glasses, is almost exclusively by grinding.

## Cross-References

- ▶ [Abrasive Machining Processes](#)
- ▶ [Contact Temperature of a Moving Solid Surface](#)
- ▶ [Cutting Fluids and Their Environmental Impact](#)
- ▶ [Cutting Tool Wear and Failure Mechanisms](#)
- ▶ [Environmentally Friendly Lubrication Issues](#)
- ▶ [Friction Modeling for Machining](#)
- ▶ [Function of Cutting Fluids and Lubricants](#)
- ▶ [Manufacturing Tribology](#)
- ▶ [Metalworking Lubricants](#)
- ▶ [Minimum Quantity Lubrication \(MQL\) During Machining of Automotive Components](#)
- ▶ [Surface Statistics and Probability Density Function](#)
- ▶ [Tribology in Minimum Quantity Lubrication \(MQL\) Cutting](#)

## References

- C. Andrew, T.D. Howes, T.R.A. Pearce, *Creep Feed Grinding* (Holt, Rinehart, and Winston, London, 1985)
- S. Kalpakjian, *Manufacturing Processes for Engineering Materials* (Addison Wesley, Reading, 1984), p. 573
- S. Malkin, C. Guo, Thermal analysis of grinding. *Ann. CIRP* **56**(2), 760–782 (2007)
- S. Malkin, C. Guo, *Grinding Technology: Theory and Applications of Machining with Abrasives*, 2nd edn. (Industrial Press, New York, 2008)
- I.D. Marinescu, M. Hitchiner, E. Uhlmann, W.B. Rowe, I. Inasaki, *Handbook of Machining with Grinding Wheels* (CRC Press/Taylor & Francis, Boca Raton, 2007)
- M.C. Shaw, *Principles of Abrasive Processing* (Oxford University Press, Oxford, 1996)

## Growth Characteristics of Large Fatigue Cracks

KWAI S. CHAN

Mechanical Engineering Division, Southwest Research Institute®, San Antonio, TX, USA

## Synonyms

[Continuum cracks](#); [LEFM \(linear-elastic fracture mechanics\) cracks](#); [Macro-cracks](#); [Macroscopic fatigue cracks](#)

## Definition

Metal fatigue is a localized damage process induced in metallic materials as the result of cyclic loadings (▶ [Fatigue](#)). The damage evolution process involves localized plastic deformation, the development of dislocation structures, and the initiation of microcracks at a variety of features that may include dislocation pileups, dislocation cells, slip bands, intrusions, extrusion, pores, inclusions, grain boundaries, notches, surface scratches, machining marks, and, among others, fretting surfaces or corrosion pits. Once formed, a microcrack may arrest or grow into a macroscopic fatigue crack. Alternately, a macroscopic fatigue crack can also be formed by coalescence of adjacent microcracks. The growth characteristics of microcracks, often referred to as small fatigue cracks, are described in another entry (▶ [Growth Characteristics of Small Fatigue Cracks](#)). The growth characteristics of a macroscopic fatigue crack, also known as a large fatigue crack, are presented in this entry. When the plastic zone of a large crack is sufficiently small that it satisfies the conditions of small-scale yielding, then linear elastic fracture mechanics (LEFM) is applicable and the fatigue crack can be referred to as an LEFM crack or a continuum crack. The latter occurs when the growth behaviors of the large crack become insensitive to the microstructure, which occurs when the crack-tip plastic zone interrogates a sufficiently large number of grains.

## Scientific Fundamentals

### Linear-Elastic Fracture Mechanics

Fracture mechanics is a field of mechanics that is concerned with the formation and growth of cracks in materials and the description of the crack behaviors in terms of an appropriate crack driving force parameter. In general, there are three modes of cracking, corresponding to opening (Mode I), in-plane shear



(Mode II), and out-of-plane shear (Mode III). For linear-elastic materials, the stress field near the tip of a large crack is given by (Kanninen and Popelar 1985)

$$\sigma_{ij} = \frac{K_l}{\sqrt{2\pi r}} f_i(\theta) \quad (1)$$

where  $K_l$  is the stress intensity factor for the  $l^{th}$  mode of cracking and  $l = I, II$ , and  $III$  for opening, in-plane shear, and out-of-plane shear mode, respectively;  $r$  is the distance measured from the crack origin;  $\theta$  is the angular position measured anticlockwise from the crack origin; and  $f_i(\theta)$  represents the appropriate angular function for the mode  $l$  crack. The stress intensity factor,  $K$ , characterizes the near-tip stress field of an elastic material since both the local stresses and the plastic zone size scales with this fracture mechanics parameter.

The stress intensity factor,  $K$ , for an elastic crack is described by the general expression given as (Kanninen and Popelar 1985; ► [Stress Intensity Factors](#))

$$K_I = F_I(a/W) \sigma_I \sqrt{\pi a} \quad (2)$$

where  $F_I$  is a boundary-correction factor,  $\sigma_I$  is the applied stress, and  $a$  is the length of the Mode  $I$  crack. The boundary-correction is a function of specimen geometry and the ratio,  $a/W$ , of crack length,  $a$ , to the width,  $W$ , of the cracked specimens.

For cyclic loading, the stress ranges,  $\Delta\sigma_{ij}$ , near the tip of a fatigue crack, is given by (Kanninen and Popelar 1985)

$$\Delta\sigma_{ij} = \frac{\Delta K_l}{\sqrt{2\pi r}} f_i(\theta) \quad (3)$$

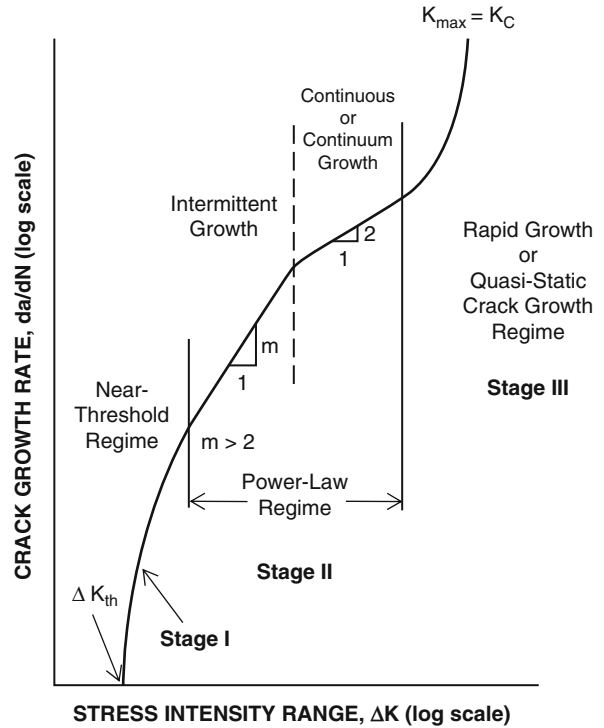
and  $\Delta K$ , the stress intensity range for the mode  $l$  crack, has the form given by (Kanninen and Popelar 1985)

$$\Delta K_I = F_I(a/W) \Delta\sigma_I \sqrt{\pi a} \quad (4)$$

where  $\Delta\sigma_I$  is the applied stress range. The boundary-correction factor is applicable for both monotonic and cyclic loading.

During cyclic loading, a monotonic plastic zone forms first at the tip of a large fatigue crack upon application of the initial applied stress. The magnitude of the monotonic plastic zone size depends on the maximum value of the stress intensity factor,  $K_{max}$ . Upon unloading to the minimum stress in the fatigue cycle, reverse yielding occurs to form a cyclic plastic zone near the crack tip. During stress cycling between the minimum stress,  $\sigma_{min}$ , to  $\sigma_{max}$ , the corresponding stress intensity factor increases from  $K_{min}$  to  $K_{max}$ , producing a stress intensity range given by

$$\Delta K = K_{max} - K_{min} \quad (5)$$



**Growth Characteristics of Large Fatigue Cracks, Fig. 1**

A schematic showing the various growth regimes in a fatigue crack growth curve of a large crack

and a stress ratio,  $R$ , which is defined as

$$R = \frac{\sigma_{min}}{\sigma_{max}} = \frac{K_{min}}{K_{max}} \quad (6)$$

for the loading cycle. The size of the cyclic plastic zone is dictated by the stress intensity factor range and the cyclic yield strength of the material.

### General Fatigue Crack Growth Characteristics

The growth kinetics of fatigue cracks are generally characterized in terms of fatigue crack growth rate,  $da/dN$ , which is a measure of the incremental increase in the crack length per fatigue cycle as a function of a fracture mechanic parameter called the stress intensity factor range,  $\Delta K$ . The fatigue crack growth rate is typically presented as a function of  $\Delta K$  in a double logarithm plot, as shown schematically in Fig. 1 (Fine et al. 1979; Chan 1993). The fatigue crack growth curve of large cracks in metals and alloys typically exhibits a sigmoidal shape, which can be divided into three characteristic regimes (Fine et al. 1979): (1) a near-threshold regime where the growth response is



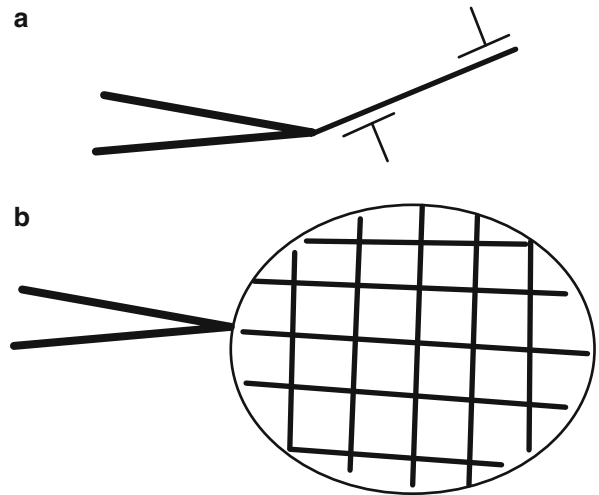
sensitive to microstructure with a threshold below which fatigue crack growth does not occur, (2) a power-law regime where microstructure has relatively small effect on the propagation rate, and (3) a rapid growth regime that is dominated by quasi-static crack extension and fracture toughness (e.g., the  $K_{IC}$  value), and is sensitive to microstructure. As shown in Fig. 1, the power-law regime may be subdivided into intermittent and continuous growth regimes (Chan 1993). For many alloys, there is considerable overlap between the intermittent and continuous growth regimes such that they become indistinguishable and appear as a single power-law regime.

### Fatigue Crack Growth Thresholds

Large cracks in metals subjected to cyclic loading generally exhibit a threshold stress intensity range,  $\Delta K_{th}$ , below which fatigue crack growth (FCG) ceases to occur. The large-crack threshold value is an important material parameter in life-prediction methods based on fracture mechanics since the fatigue crack growth life, most of which is spent in the near-threshold regime, is very sensitive to the critical stress intensity range where crack extension begins. The  $\Delta K_{th}$  value depends on a large number of factors, including microstructure, loading history, temperature, frequency, crack size, and environment (Fine et al. 1979; Chan 2004). Furthermore, the value of  $\Delta K_{th}$  shows substantial variations due to material variability and testing methods. For example, the  $\Delta K_{th}$  of structural alloys shows large variations with the stress ratio,  $R$ , or the mean stress (Chan 2004).

Studies of fatigue mechanisms in the near-threshold regime have revealed the presence of extrinsic and intrinsic processes (Chan 2004) that can alter significantly the threshold value at which fatigue crack growth ceases to occur. As illustrated in Figs. 2a, b, intrinsic fatigue mechanisms are those that usually operate at or ahead of the crack tip and reflect the material's resistance to cyclic deformation and fatigue failure in the crack-tip region; for example, they include the emission of dislocations from the crack tip or the to-and-fro motion of dislocations in a cellular network formed ahead of the crack tip. In contrast, extrinsic fatigue mechanisms, which are summarized in Figs. 3a–d (Ritchie et al. 1986), are those that affect the driving force (i.e.,  $\Delta K$ ); they usually operate in the crack wake and include the crack closure mechanisms due to plastic wake (Elber 1970), crack deflection and branching (Suresh 1985), fracture surface asperity and roughness (Suresh and Ritchie 1982), ligament toughening, and oxide wedging (Suresh and Ritchie 1983).

The experimental evidence of an intrinsic and an extrinsic fatigue crack growth threshold is presented in

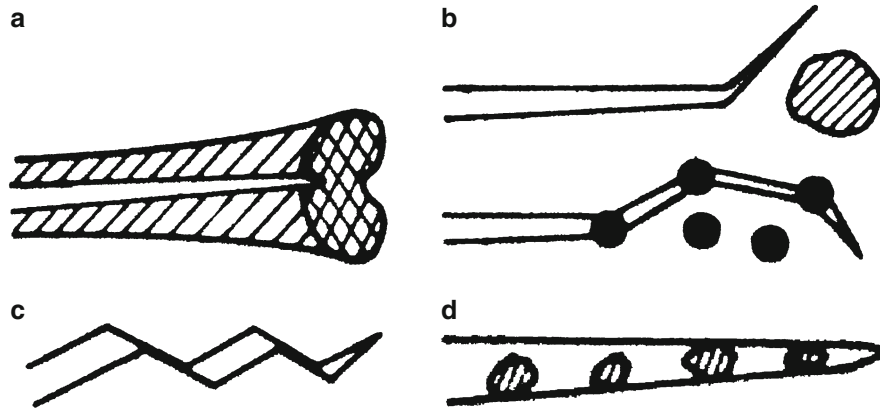


**Growth Characteristics of Large Fatigue Cracks, Fig. 2** Schematics of intrinsic fatigue threshold mechanisms: (a) dislocation dipole or pile-up in planar slip materials, and (b) dislocation cellular network and ladder structure in cell-forming materials (Chan 2004)

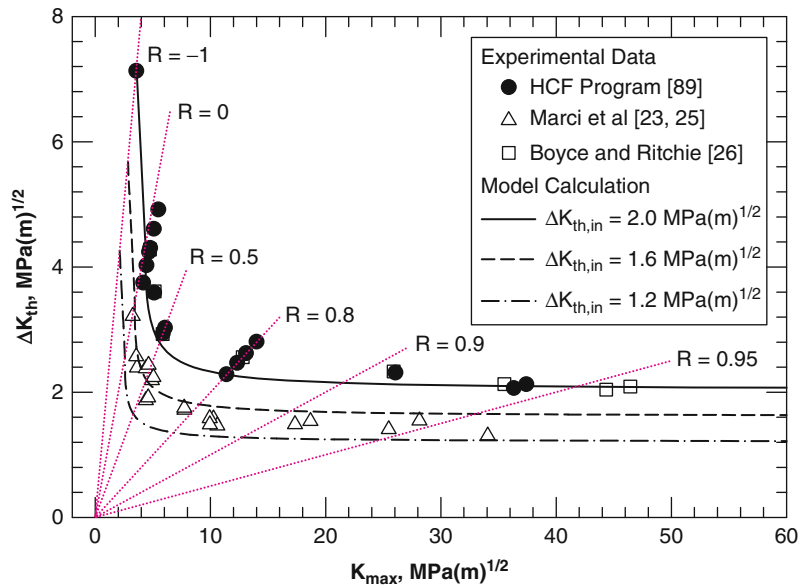
Fig. 4, which shows a plot of the  $\Delta K_{th}$  of large cracks as a function of the maximum stress intensity factor,  $K_{max}$ , in a fatigue cycle for Ti-6Al-4 V. Two regimes are evident in Fig. 4. At high  $K_{max}$ , the value of  $\Delta K_{th}$  is a constant that is essentially independent of  $K_{max}$  or the  $R$  ratio. The lack of dependence of  $\Delta K_{th}$  on  $K_{max}$  is an indication of an intrinsic FCG threshold that is independent of loading history. In contrast, the  $\Delta K_{th}$  value increases with decreasing  $K_{max}$  at low  $K_{max}$  values. The increase of  $\Delta K_{th}$  with  $K_{max}$  is the result of crack closure, which becomes prominent when  $K_{max}$  is less than that required to fully open the crack tip. The dependence of  $\Delta K_{th}$  on  $K_{max}$  is an indication of an extrinsic FCG threshold and the presence of crack-tip shielding mechanisms that lower the near-tip crack driving force (i.e., stress intensity range). Crack shielding mechanisms are often referred to as crack closure mechanisms, which include plastic stretch in the crack wake, crack deflection and branching, surface asperity and roughness-induced closure, ligament toughening, and oxide wedging.

### Intrinsic Threshold

The intrinsic threshold arises from damages induced by dislocation processes at the crack tip. One of the earliest fatigue crack growth intrinsic threshold models is that of Weertman et al. 1982, who considered that the onset of fatigue crack growth is dictated by



**Growth Characteristics of Large Fatigue Cracks, Fig. 3** Schematics of crack closure mechanisms: (a) plasticity-induced crack closure, (b) crack deflection and bridging, (c) roughness-induced crack closure, and (d) oxide-induced crack closure (Ritchie and Wu, 1987)



**Growth Characteristics of Large Fatigue Cracks, Fig. 4** Measured and theoretical  $\Delta K_{th}$  for Ti-6Al-4 V as a function of  $K_{max}$  from (Chan 2004) (Sources of experimental data are shown in Chan (2004))

the emission of a dislocation from the crack tip governed by (Weertman et al. 1982)

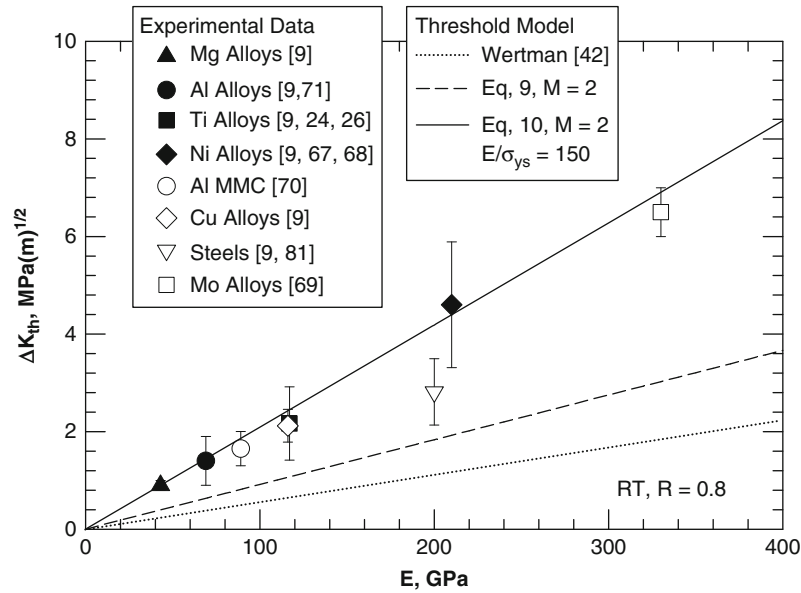
$$K_{th} = \frac{1}{3} E b^{1/2} \quad (7)$$

where  $K_{th}$  is the threshold stress intensity for emitting a dislocation from the crack tip,  $E$  is young's modulus, and  $b$  is the magnitude of the Burgers vector. Weertman made a comparison of (7) against experimental data for low  $R$  ratios and found (7) to underpredict the fatigue crack growth thresholds by a factor of 2–5 for

most metals (Weertman et al. 1982). Other damage processes that have been considered include to-and-fro dislocation motion or cyclic slip at a characteristic distance,  $r_s$ , ahead of the crack tip, and the development of a dislocation cell structure by cyclic slip in the near tip region.

The various threshold models in the literature have been generalized to give (Chan 2004)

$$\Delta K_{th,in} = \frac{\sqrt{12}M}{8\sqrt{\pi}(1+\nu)} \left(\frac{b}{h}\right)^{1/2} E b^{1/2} \quad (8)$$



**Growth Characteristics of Large Fatigue Cracks, Fig. 5** Plot of intrinsic fatigue crack growth threshold,  $\Delta K_{th,in}$ , at  $R = 0.8$  versus Young's modulus  $E$ , compared against pertinent model computations (Chan 2004) (Sources of experimental data are shown in Chan (2004))

for the intrinsic threshold of planar slip materials, where  $\nu$  is Poisson's ratio,  $M$  is the Taylor factor, and  $h$  is the slip band width. The values of  $b/h$  ranges from 1.1547 for fcc metals to 1.414 for bcc metals and  $\nu = 1/3$  for isotropic solids. Thus,

$$\Delta K_{th,in} = 0.2738MEb^{1/2} \quad (9)$$

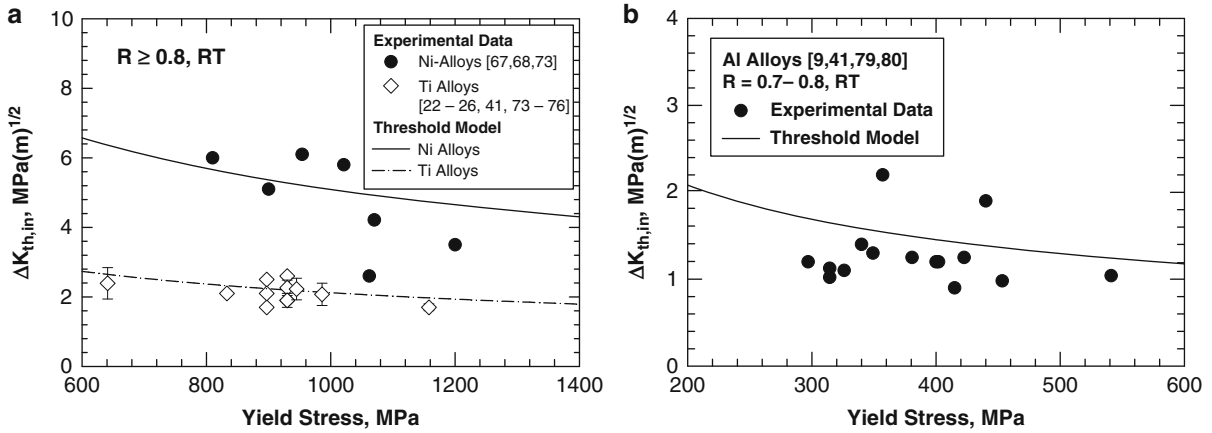
for materials that exhibits planar slip. For polycrystalline materials, the lowest value of the Taylor factor,  $M$ , is 2. Hence, the lowest bound fatigue crack growth threshold is  $\Delta K_{th,in} = 0.55Eb^{1/2}$ . For cell-forming materials (Chan 2004),

$$\Delta K_{th,in} = 0.1 \left[ \frac{ME}{\sigma_{ys}} \right]^{1/2} Eb^{1/2} \quad (10)$$

for an intrinsic threshold of moving a single dislocation to-and-fro between the walls of a dislocation cellular structure in wavy slip materials.

The intrinsic threshold models, (9 and 10), indicate that the  $\Delta K_{th,in}$  increases with increasing Young's modulus for both planar slip and cell-forming materials. In the latter case,  $\Delta K_{th,in}$  also depends mildly on the yield strength and the Taylor factor. Since  $\sigma_{ys} = M\tau_{ys}$ ,  $\Delta K_{th,in} \propto \tau_{ys}$ , with the critical resolved shear stress,  $\tau_{ys}$ , being the only microstructural parameter in (10). Figure 5 presents a compilation of  $\Delta K_{th,in}$  as a function of Young's modulus for various alloys. Experimental data of  $\Delta K_{th,in}$  and  $E$  shown in Fig. 5 are compiled from the literature

(Chan 2004) and magnitude of the Burgers vector,  $b$ , is taken to be  $2.8 \times 10^{-10}$  m. The  $\Delta K_{th}$  values at  $R = 0.8$  were chosen as the intrinsic thresholds because crack closure is expected to be negligible at this level of  $R$  value. The data in Fig. 5 show that  $\Delta K_{th,in}$  indeed relates linearly with  $E$  to within the experimental scatter. Fig. 5 also shows a comparison of experimental data of  $\Delta K_{th,in}$  against the Weertman model, (7), and the planar slip (dislocation dipole) model, (9). In both cases, the models underpredicted the intrinsic thresholds as shown in Fig. 5. For comparison, Fig. 5 shows a comparison of the experimental  $\Delta K_{th,in}$  data against the theoretical threshold model, (10), for cell-forming materials at  $E/\sigma_{ys} = 150$ , which is in good agreement with experimental data. According to (10),  $\Delta K_{th,in}$  increases with decreasing yield stress to the  $-1/2$  power. This theoretical prediction is evaluated against experimental data compiled for Ni-alloys, Ti-alloys, and Al-alloys, in Fig. 6 (Chan 2004). In general, model predictions based on (10) are in good agreement with the experimental data with the exception of steels. Figure 6 shows that the dependence of  $\Delta K_{th,in}$  on yield strength is generally mild, as predicted by the theory. Another implicit prediction of (9 and 10) is that  $\Delta K_{th,in}$  does not depend explicitly on the microstructural size scale, which appears to be borne out by experimental observations since  $\Delta K_{th}$  data of various grain sizes are included in Fig. 6.



**Growth Characteristics of Large Fatigue Cracks, Fig. 6** Comparison of observed and computed intrinsic threshold,  $\Delta K_{th,in}$ , as a function of yield stress for  $R = 0.8$ : (a) Ti-alloys and Ni-alloys, and (b) Al-alloys (Chan 2004) (Sources of experimental data are shown in Chan (2004))

### Extrinsic Threshold

Many crack closure models have been proposed in the literature, including those for treating plasticity-induced crack closure, crack deflection, fracture-surface roughness, and oxide-induced crack closure. Essential features of these various crack closure mechanisms are summarized and integrated into a set of analytical equations for computing the contributions of extrinsic mechanisms to the large-crack fatigue crack growth threshold.

The combination of plasticity-induced closure with crack deflection, asperity-induced, and oxide-induced crack closure leads one to (Chan 2004)

$$\Delta K_{eff} = U \Delta K \quad (11)$$

where  $U$  incorporates all relevant crack closure mechanisms. Since  $U_i$  ( $i = P, D, R$ , and  $O$ ) for individual closure mechanisms are defined as a fraction of  $\Delta K_{op}$  to the applied  $\Delta K$ , the proper approach to incorporate multiple crack closure mechanisms is to take the product of individual  $U_i$  by considering that the open portion of the fatigue crack is acted upon successively by individual closure effects. On this basis,  $U$  is given by (Chan 2004)

$$U = U_P U_D U_R U_O \quad (12)$$

for all four closure mechanisms are considered. Furthermore, (11) can be equated to the intrinsic threshold,  $\Delta K_{th,in}$ , to give (Chan 2004)

$$\Delta K_{eff} = U \Delta K = \Delta K_{th,in} \quad (13)$$

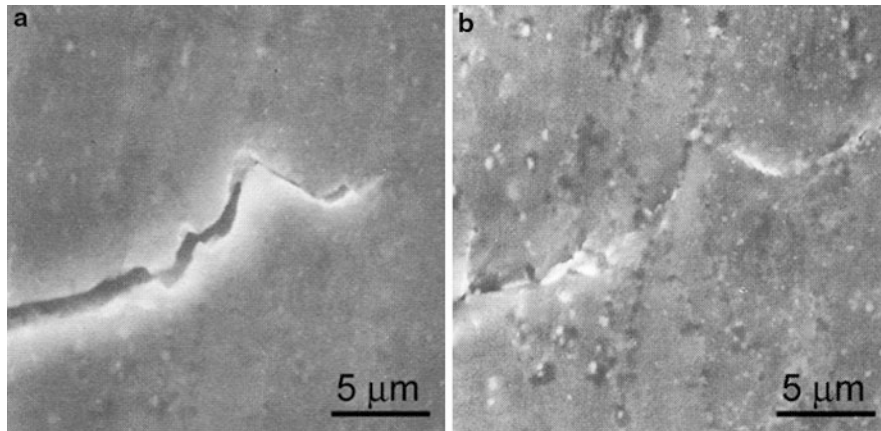
at the fatigue threshold ( $\Delta K = \Delta K_{th}$ ). Hence, (12) can be rearranged to obtain

$$\Delta K_{th} = U^{-1} \Delta K_{th,in} = [U_P U_D U_R U_O]^{-1} \Delta K_{th,in} \quad (14)$$

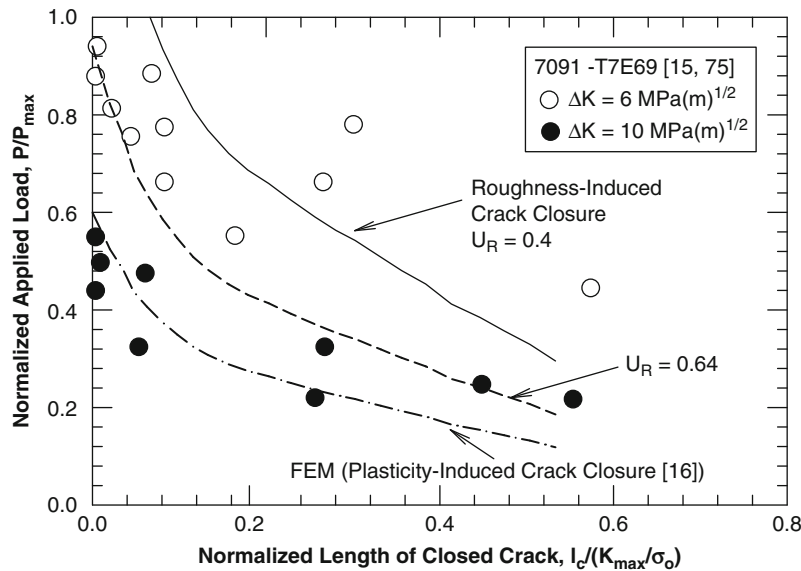
as the fatigue crack growth threshold for large cracks in structural alloys exhibiting both intrinsic and extrinsic fatigue mechanisms. The expressions for  $U_P$ ,  $U_D$ ,  $U_R$ , and  $U_O$  are given in Chan (2004).

Evidence of extrinsic crack closure mechanisms is illustrated in Fig. 7, which shows fatigue crack closure in 7091-T7E69 for  $\Delta K = 8 \text{ MPa}\sqrt{\text{m}}$  at  $R = 0.16$  from Davidson and Lankford (1984). In Fig. 7a, the fatigue crack is seen to fully open at the maximum load. In contrast, the fatigue crack is fully closed near the tip at the minimum load, Fig. 7b. In situ observations of the fatigue crack growth process also indicated that crack path exhibited a high degree of tortuosity with mixed Mode I and II crack opening displacements (Davidson and Lankford 1984). Only a fraction of measured Mode II CTOD was required to cause the observed roughness-induced crack closure. During loading, the crack surface peeled open at distances far behind the crack tip. The opening gradually moved toward the crack tip upon increasing  $K$  levels. In general, the crack tip was closed at a large portion of the fatigue cycle at low  $R$  (e.g.,  $R \leq 0.2$ ) and cyclic slip could not occur ahead of the crack tip until the crack tip was completely open at  $K \geq K_{cl}$ .

The presence of concurrent crack closure mechanisms can be demonstrated by comparing experimental and theoretical crack opening loads for individual closure mechanisms. Previously, McClung and Davidson (1991) compared experimentally measured crack opening loads for the 7091 Al alloy against those predicted based on plasticity-induced crack closure and finite-element method. As shown in Fig. 8, their results indicated that the predicted crack opening loads were substantially lower



**Growth Characteristics of Large Fatigue Cracks, Fig. 7** Near-tip crack opening behavior of a fatigue crack in 7091-T7E69 subject to  $\Delta K = 8 \text{ MPa}\sqrt{\text{m}}$  at  $R = 0.16$ : (a) crack opening at  $K_{\max}$ , and (b) crack closure at  $K_{\min}$ . (Davidson and Lankford 1984)

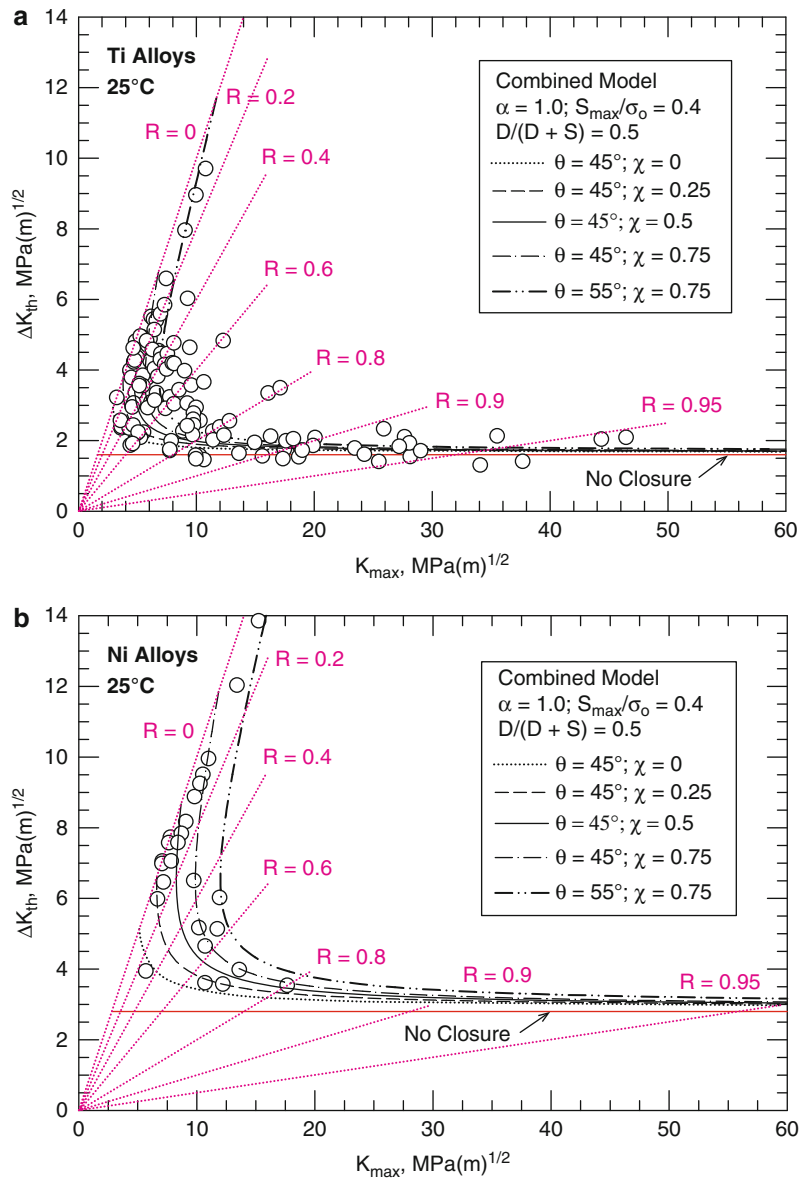


**Growth Characteristics of Large Fatigue Cracks, Fig. 8** A comparison of computed and measured crack opening loads suggest the presence of concurrent plasticity- and roughness-induced crack closure mechanisms in 7091 Al (Chan 2004) (Sources of experimental data are shown in Chan (2004))

than experimental observations, meaning crack closure in the specimens was larger than those predicted by plasticity-induced crack closure. The crack opening load results were used by Chan to deduce the various levels of roughness-induced crack closure. The results are presented in Fig. 8, which shows that the presence of concurrent roughness-induced crack closure would increase the normalized crack opening load from 0.6 to

0.9 when  $U_R = 0.4$ –0.64. This result indicates that roughness-induced crack closure could be as large as that of plasticity-induced crack closure.

The influence of  $K_{\max}$  on  $\Delta K_{\text{th}}$  is presented in Fig. 9 for Ti-6Al-4V. Lines of constant  $R$  ratios are also superimposed in Fig. 9. In general,  $\Delta K_{\text{th}}$  is relatively constant and insensitive to  $K_{\max}$  when  $R > 0.8$ . On the other hand,  $\Delta K_{\text{th}}$  increases rapidly with decreasing  $K_{\max}$  when



**Growth Characteristics of Large Fatigue Cracks, Fig. 9** Measured  $\Delta K_{th}$  versus  $K_{max}$  compared against crack closure models: (a) Ti-alloys and (b) Ni-alloys (Chan 2004) (Sources of experimental data are shown in Chan (2004))

$R < 0.8$  and results in variations of  $\Delta K_{th}$  value at a given value of  $K_{max}$ . For example, at  $K_{max} = 10 \text{ MPa}\sqrt{\text{m}}$ ,  $\Delta K_{th}$  ranges from 2 to 10  $\text{MPa}\sqrt{\text{m}}$ . Computed curves based on various crack closure mechanisms are compared against experimental data in Fig. 9. In all cases, the computed  $\Delta K_{th}$  values at  $R > 0.8$  are essentially independent of the R ratio or  $K_{max}$ , but increase rapidly almost vertically when  $K_{max}$  is lower than an apparent critical  $K_{max}$  value, which is

about  $10 \text{ MPa}(\text{m})^{1/2}$ . For most cases, the computed  $\Delta K_{th}$  goes up almost vertically at the apparent critical  $K_{max}$  ( $\approx 10 \text{ MPa}(\text{m})^{1/2}$ ), but sometimes it can slant backward and exhibit a positive slope, as shown by the two dot-dashed lines in the figure. The  $K_{max}$  value where  $\Delta K_{th}$  goes up vertically at small changes of  $K_{max}$  is also somewhat dependent on the crack closure model and the amount of crack closure, as evidenced by the computed curves in Fig. 9.



The shape of the  $\Delta K_{th} - K_{max}$  curves predicted by the theoretical threshold model can be readily explained by recognizing that, in general (Chan 2004),

$$\Delta K_{th}(R, K_{op}/K_{max}) = \left[ \frac{1-R}{1 - \frac{K_{op}}{K_{max}}} \right] \Delta K_{th,in} \quad (15)$$

where  $\Delta K_{th}(R, K_{op}/K_{max})$  denotes that  $\Delta K_{th}$  is a function of the  $R$  ratio as well as the  $K_{op}/K_{max}$  ratio. At high  $R$  ratios ( $R > 0.8$ ), the crack opens at  $K_{min}$ . Thus,  $K_{op} = K_{min}$ ,  $R = K_{op}/K_{max}$ , and  $\Delta K_{th} = \Delta K_{th,in}$ . This results in a horizontal line in a plot  $\Delta K_{th}$  versus  $K_{max}$ . In the case of  $K_{max} \leq K_{op}$ , the crack is closed during the entire fatigue cycle; it would not propagate and exhibits a threshold-like behavior. In contrast, if a fatigue crack is closed during most of the fatigue cycle, then  $K_{op}/K_{max}$  approaches unity and the ratio of  $(1-R)/(1-K_{op}/K_{max})$  becomes large. This behavior gives rise to the vertical boundary in the plot of  $\Delta K_{th}$  versus  $K_{max}$  at low  $R$  values. Furthermore,  $\Delta K_{th}$  decreases rapidly with increasing  $K_{max}$  and it gives rise to the appearance that a critical  $K_{max}$  must be exceeded. In reality, this critical  $K_{max}$  may simply be the stress intensity level required to keep the fatigue crack tip open during at least a portion of the fatigue cycle so that  $\Delta K_{eff} \geq \Delta K_{th,in}$ .

### Near-Threshold Regime

At  $\Delta K$  just above the large-crack FCG threshold, the fatigue crack growth rate,  $da/dN$ , increases rapidly with only a slight increase in  $\Delta K$ . Referred to as the near-threshold regime, this regime is characterized by Stage I crack growth where the fatigue crack propagates along a slip plane and exhibits a strong dependence in the form of a steep slope in the relation between  $\log(da/dN)$  and  $\log(\Delta K)$ , for  $\Delta K$  values just above the  $\Delta K_{th}$ .

The criterion for slip-band cracking is the range of resolved shear stress on the slip plane. Consequently, Stage I crack growth in the near-threshold regime occurs on slip planes aligned with the maximum shear stress ranges of the crack-tip stress field. Stage I crack is often a mixed-mode crack with a combination of Mode I, Mode II, and Mode III components. The Mode II and Mode III components provide the driving force for cyclic slip coplanar with the crack plane and cracking along the crystallographic slip plane, while the Mode I component provides the driving force for out-of-plane slip and opening up the crack surfaces. Coplanar slip ahead of a Stage I crack tends to limit out-of-plane slip and allows a buildup of a high normal stress acting on the slip plane (Koss and Chan, 1980). As a result, a Stage I crack often exhibits a cleavage-like fracture appearance, despite

the fact that cracking occurs along localized slip bands (Koss and Chan, 1980). Because the opening mode is generally small, Stage I fatigue cracking is susceptible to crack closure mechanisms such as crack deflection and fracture surface asperities when the slip band crack meanders to propagate along two or more parallel slip planes. The meandering crack path can result in mismatching crack planes as the crack traverses from one grain to a neighboring grain, resulting in ligament toughening and crack-tip shielding. The consequence is that Stage I crack growth can lead to a very steep slope in the  $da/dN$  versus  $\Delta K$  curve.

One of the most important features of the near-threshold fatigue regime is its strong dependence on mean stress (Fine et al. 1979). As indicated earlier in the section on fatigue thresholds, the strong mean stress dependence is directly related to the presence of crack closure mechanisms that reduce the crack-tip stress intensity range (i.e., crack-tip shielding) below the nominal value provided by the external loads. A high mean stress or  $R$  ratio tends to open up the crack, break up bridging ligaments in the crack wake, and reduce the effectiveness of the crack closure or crack-tip shielding mechanism; thereby, the apparent  $\Delta K_{th}$  is markedly reduced to its intrinsic value.

The fracture morphology in the near-threshold regime is well characterized and is very microstructurally sensitive due to the presence of isolated planar transgranular or intergranular facets within a flat ductile transgranular mode for fatigue crack growth in air at ambient temperature (Fine et al. 1979). In polycrystalline materials, Stage I fatigue facets are intergranular in ferritic steels and transgranular in stainless steels, titanium alloys, and Ni-based superalloys. They appear to be environmentally induced, since they are largely suppressed in vacuum. The isolated planar facets are created by cracking along slip bands, while intergranular facets are formed by cracking along grain boundaries. The facets are sometimes mixed with grains that exhibit fatigue striations, which are formed by crack-tip blunting via alternating slip on two inclined slip planes. This particular fatigue crack growth mechanism, which is responsible for Stage II growth, will be discussed in greater detail in the power-law growth regime.

Extended Stage I crack growth occurs in single crystal alloys, including Ti alloys and Ni-based superalloys. The FCG growth of large cracks in single crystal alloys exhibit a near-threshold regime, a power-law regime, and the rapid growth regime (i.e., the same characteristics as those observed in polycrystalline materials). Thus, Stage I growth is not limited to the near-threshold regime and

vice versa. Similarly, Stage II growth is not limited to the power-law regime and vice versa.

### Power-Law Growth Regime

The power-law growth regime was first investigated by Paris, who pioneered the use linear-elastic fracture mechanics for treating fatigue crack growth in metals by correlating the incremental crack extension per fatigue cycle,  $da/dN$ , to the stress intensity range,  $\Delta K$ , in a double logarithm plot. He showed that a power-law relation, known as the Paris's relation, exists between  $\log(da/dN)$  and  $\log(\Delta K)$  as given by (Paris and Erdogan 1963)

$$\frac{da}{dN} = C \Delta K^m \quad (16)$$

where  $C$  and  $m$  are empirical constants. Modifications of the Paris equation to treat stress ratio effects or the presence of a growth threshold are also available in the literature. Most of these models are empirical and provide little information about the effects of microstructural size scale on the crack growth kinetics.

The slope in the power-law regime is typically 2–8. In most materials, the power-law regime can be further divided into an intermittent growth regime with a slope of 4 where fatigue crack growth does not occur at every cycle and extends only after damage accumulation within the crack tip cyclic plastic zone exceeds a critical value. In contrast, fatigue crack growth occurs during every cycle at high  $\Delta K$ . In this continuum growth regime, the crack growth rate, which is in the range of  $10^{-6}$  to  $10^{-3}$  m/cycle, is dominated by the cyclic crack-tip opening displacement,  $\Delta CTOD$ , and a slope of 2 in the power-law regime. A wider range of slope ( $m$ ) is observed in the power-law regime because of the presence of crack closure during fatigue crack growth at low mean stresses or  $R$  ratios ( $R < 0.5$ ). The continuum growth regime usually occurs at high  $\Delta K$  levels, where rapid crack growth may intervene and obscure the continuum crack growth regime.

### Rapid Growth Regime

At high  $\Delta K$ , the crack propagates rapidly when the maximum stress intensity factor,  $K_{max}$ , approaches the critical stress intensity factor,  $K_{IC}$ , for unstable fracture of a Mode I crack. The fracture mode is a combination of fatigue crack growth and quasi-static fracture, which is influenced by microstructure, mean stress, and specimen thickness. In this regime, the fatigue striations are generally coarse. The striations are often intermixed with dimples, tear ridges, cleavage-like facets, or grain boundary facets, depending on materials and microstructures. The critical  $K_C$  value at fatigue fracture can exceed the  $K_{IC}$  value

because the plane strain condition may not be met. For brittle materials, the occurrence of concurrent fatigue crack growth and quasi-static fracture results in the dependence of  $da/dN$  on both  $\Delta K$  and  $K_{max}$ .

### Fatigue Crack Growth Models

The most common FCG model is the power-law equation, (16), proposed by Paris. This model treats only Stage II growth and ignores the presence of the growth threshold, Stage I and Stage III growth. Nonetheless, it is used extensively for estimating the FCG life because of its simplicity and both  $\Delta K_{th}$  and  $K_{IC}$  can be implemented rather easily.

The Paris power-law equation has been modified by Newman to include an effective stress intensity range, which is obtained by subtracting a crack closure term due to plasticity-induced crack closure from the applied stress intensity range. A Dugdale-type strip model was formulated and implemented into a software package called NASGRO to compute the fatigue crack growth rate of large cracks for various crack geometries in structure. The NASGRO package has been commercialized and updated regularly. The current NASGRO equation describes all three regions of the large-crack FCG curve and is given by (Newman et al. 1981)

$$\frac{da}{dN} = C \left( \frac{(1-f)\Delta K}{1-R} \right)^n \frac{\left( 1 - \frac{\Delta K_{th}}{\Delta K} \right)^p}{\left( 1 - \frac{K_C}{K_{max}} \right)^q} \quad (17)$$

where  $C$  and  $n$  are empirical constants for the power-law region. The parameter  $p$  is an empirical constant describing the large-crack threshold region,  $\Delta K_{th}$ , and the parameter  $q$  describes the fast fracture region where the maximum  $K$ ,  $K_{max}$ , approaches the critical stress intensity at fracture,  $K_C$ . The parameter  $f$  is the ratio of  $K_{op}/K_{max}$  where  $K_{op}$  is the stress intensity factor at which the crack tip is fully open. The presence of a compressive residual increases  $K_{op}$  and thus reduces the effective  $\Delta K$ , which is the difference between  $K_{max}$  and  $K_{op}$ . The value of  $f$  is computed in NASGRO using the Newman crack closure model (1981). Both (16) and (17) are phenomenological and all the material constants in the FCG expressions must be determined empirically from experimental data.

Chan and Enright have developed a physics-based fatigue crack growth model to treat the three stages of fatigue crack growth, which has the form given by (Chan and Enright 2006)

$$\frac{da}{dN} = \frac{2s\zeta^{n_2/2}}{[E(2s)^{1/2}]^{n_2}} \left[ \Delta K_T^{n_1-n_2} \Delta K_{eff}^{-n_1} + \Delta K_{eff}^{-n_2} - [(1-R)K_C]^{-n_2} \right]^{-1} \quad (18)$$

with

$$\xi = \frac{Es}{4\sigma'_y \epsilon'_f d} \quad (19)$$

$$d = d_o \left( \frac{D}{D_o} \right)^\gamma \quad (20)$$

and

$$\Delta K_T = \frac{\Delta \sigma'_y}{\Delta \sigma_e} \Delta K_{th} \quad (21)$$

where the fatigue crack growth rate,  $da/dN$ , is expressed in terms of an effective stress intensity range,  $\Delta K_{eff}$ , which is the difference between the applied stress intensity range,  $\Delta K$ , and the stress intensity range at crack closure,  $\Delta K_{cl}$ .  $\xi$  is a dimensionless normalizing parameter defined in terms of the Young's modulus ( $E$ ), dislocation cell size ( $s$ ), cyclic yield stress ( $\sigma'_y$ ), fatigue ductility coefficient ( $\epsilon'_f$ ), and the dislocation barrier spacing ( $d$ ). The dislocation barrier spacing is taken to be a function of the grain size  $D$  as described by (20), where  $d_o$ ,  $D_o$ , and  $\gamma$  are material constants. The stress intensity range at the Stage I to Stage II transition,  $\Delta K_T$ , is a function of the fatigue limit ( $\Delta \sigma_e$ ) (► **Fatigue Limit**), cyclic yield stress range and the large-crack FCG threshold ( $\Delta K_{th}$ );  $n_1$  and  $n_2$  are the Stage I and Stage II exponents, respectively.  $R$  is the stress ratio and  $K_{IC}$  is the fracture toughness. The effective stress intensity range,  $\Delta K_{eff}$ , is obtained from (11 and 12) (Chan 2004).

Analytical expressions for computing individual  $U_i$  terms, which are summarized in a paper by Chan (Chan 2004), include Newman's formulation for treating plasticity-induced crack closure (Newman et al. 1981), Suresh's formulation for crack deflection (Suresh 1985) and roughness-induced crack closure (Suresh and Ritchie 1983), as well as an oxide-wedging model (Suresh and Ritchie 1983).

## Key Applications

The existence of crack-like flaws cannot be precluded in any engineering structure. These crack-like flaws can grow in engineering structure during service due to the presence of cyclic loads. These fatigue cracks can be detected by inspection using nondestructive evaluation (NDE) methods. An engineering concept, called damage tolerance approach, relies on the use of fracture mechanics and fatigue crack growth equations to design crack-resistant structures. This approach is commonly used in conjunction with NDE techniques to assess the reliability of a structure against fatigue crack growth failure.

In general, fatigue life can generally be considered to consist of two parts, which are crack initiation and crack

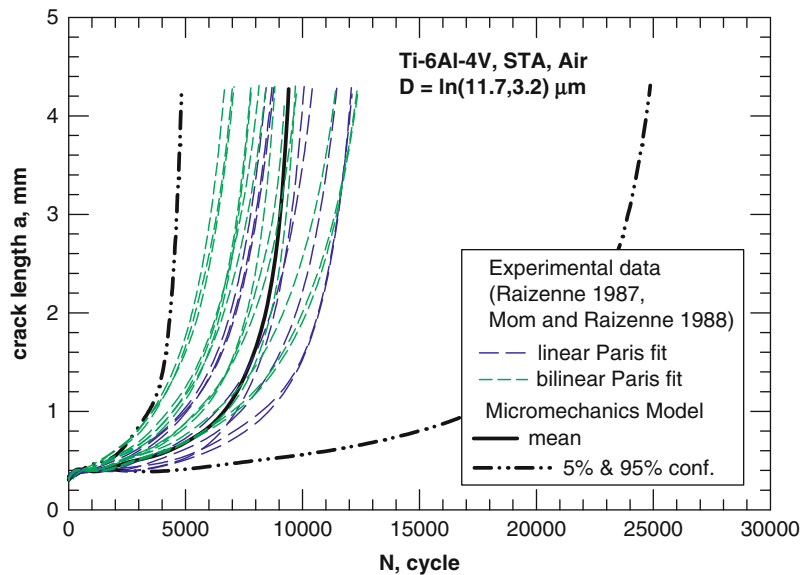
propagation lives. The crack propagation life,  $N_f$ , can be predicted by integrating the  $da/dN$  equation, leading to

$$N_f = \int_{a_i}^{a_f} \left[ \frac{da}{dN} \right]^{-1} da \quad (22)$$

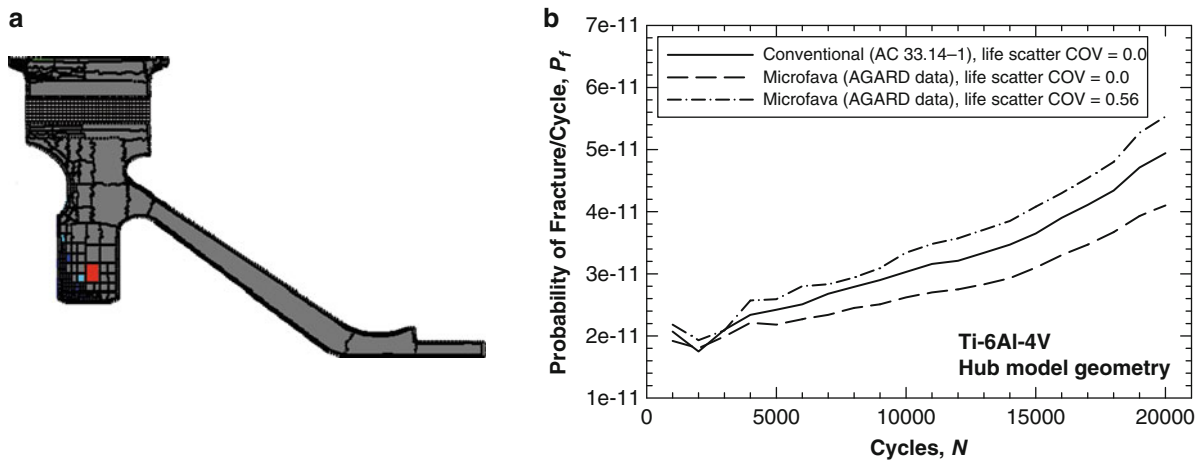
where  $a_i$  and  $a_f$  are the initial and final crack length, respectively, and  $a_f$  is defined as the critical crack length at fracture. The fatigue crack growth (FCG) life,  $N_f$ , is most sensitive to the initial crack size,  $a_i$  [1], but also depends on microstructure, which can alter intrinsic material properties as well as extrinsic properties originating from crack shielding mechanisms. To apply (22), a fatigue crack growth law must be defined for  $da/dN$  (e.g., via (16), (17), or (18)). The initial crack size,  $a_i$ , must be determined by inspection via NDE technique, but is generally assumed due to the lack of measurements or inadequate resolutions in the crack length measurements. The critical crack length is typically evaluated from the critical stress intensity factor,  $K_{IC}$ , or the fracture toughness.

As an illustration, computational and experimental  $da/dN$  values were applied to the life prediction of a 10 cm × 10 cm Ti-6Al-4 V plate. A uniform stress range of 600 MPa was applied to the plate at a stress ratio,  $R$ , of 0.1 and an initial crack (0.035 cm × 0.035 cm) was placed at one of the corners. Crack growth life was computed using probabilistic fracture mechanics software (DARWIN<sup>®</sup>) using: (1) tabular  $da/dN$  values based on (18), and (2)  $da/dN$  slope and intercept constants based on linear (Stage II) and bilinear (Stage I and Stage II) curve fits of the Paris expression, (16), to experimental data of Ti-6Al-4 V from the literature (Chan and Enright 2006). The results are shown in Fig. 10. It can be observed that all of the predicted crack growth life values based on the experimental  $da/dN$  data fall within the 5 % and 95 % confidence limits based on life values associated with the computational model based on (18).

Another illustration involves the use of (18) for predicting the fatigue crack growth life of a Ti-alloy rotor made of Ti-6Al-4 V is presented. The initial crack size was described by the distribution of the hard alpha particle size and the FCG relation given by (18) was used. The FCG life computations were performed via DARWIN<sup>®</sup> to assess the influence of variations of  $da/dN$  due to microstructural variations on the risk of fracture of the Ti rotor. Figure 11a shows the rotor disk design divided into 221 zones after 5 zone refinements. The critical region in the hub is shown as red in Fig. 11a. The predicted probability of fracture computed on the basis of conventional (experimental  $da/dN$  data) material input and (18) are compared in Fig. 11b. In Fig. 11b, the conventional data input did not consider the coefficient of variation



**Growth Characteristics of Large Fatigue Cracks, Fig. 10** Comparison of predicted crack growth life values based on microstructure-based computational model and experimental data  $da/dN$  results for  $R = 0.1$  at  $24^\circ\text{C}$  (Chan and Enright 2006)



**Growth Characteristics of Large Fatigue Cracks, Fig. 11** Analysis of a Ti rotor design using microstructure-based large-crack FCG model, (18) and DARWIN: (a) hub geometry with the critical region in red, and (b) computed probability of fracture (Chan and Enright 2006)

(COV) of  $da/dN$ . Without considering the COV due to microstructural variations, the DARWIN<sup>®</sup> prediction using (18) is slightly less conservative than that using the conventional input. Figure 11b also shows the results for which COV of  $da/dN$  due to microstructural variation is added to the mean  $da/dN$  allows one to quantify the increase in the probability of fracture due to  $da/dN$  variation. As shown in Fig. 11b, the predicted probability of fracture is increased slightly above the conventional

method line when the COV predicted from microstructural variations is incorporated into the large-crack life assessment analysis.

## Cross-References

- [Fatigue](#)
- [Fatigue Limit](#)
- [Growth Characteristics of Small Fatigue Cracks](#)
- [Stress Intensity Factors](#)

## References

- K.S. Chan, Scaling laws for fatigue crack growth of large cracks in steels. *Metall. Trans. A* **24A**, 2473 (1993)
- K.S. Chan, Variability of large-crack fatigue crack growth thresholds in structural alloys. *Metall. Mater. Trans. A* **35A**, 3721 (2004)
- K.S. Chan, M.P. Enright, A probabilistic micromechanical code for predicting fatigue life variability: model development and application. *J. Eng. Gas Turbines Power* **129**, 889 (2006)
- D.L. Davidson, J. Lankford, Fatigue crack tip mechanics of a powder metallurgy aluminum alloy in vacuum and humid air. *Fatigue Eng. Mater. Struct.* **7**, 29 (1984)
- W. Elber, Fatigue crack closure under cyclic tension. *Eng. Fract. Mech.* **2**, 37 (1970)
- M.E. Fine, R.O. Ritchie, *Fatigue-Crack Initiation and Near-Threshold Crack Growth*, ed. by M. Meshii. Fatigue and Microstructure, ASM (Metals Park, OH, 1979), p. 245
- M.F. Kanninen, C.H. Popelar, *Advanced Fracture Mechanics* (Oxford University Press, New York, 1985)
- D.A. Koss, K.S. Chan, Fracture along planar slip bands. *Acta Metall.* **28**, 1245 (1980)
- R.C. McClung, D.L. Davidson, High resolution numerical and experimental studies of fatigue cracks. *Eng. Fract. Mech.* **39**, 113 (1991)
- J.C. Newman, Jr, in *Methods and Models for Predicting Fatigue Crack Growth Under Random Loading*, ASTM STP-748, ed. by J.B. Chang, C.M. Hudson, ASTM (Philadelphia, PA, 1981), p. 53
- P.C. Paris, F. Erdogan, A critical analysis of crack propagation laws. *J. Basic Eng. Trans., ASME, Ser. D* **85**, 528 (1963)
- R.O. Ritchie, W. Yu, *Short Crack Effects in Fatigue: A Consequence of Crack Tip Shielding*, ed. by R.O. Ritchie, J. Lankford. Small Fatigue Cracks, TMS (Warrendale, PA, 1986), p. 167
- S. Suresh, Fatigue crack deflection and fracture surface contact: micromechanical models. *Metall. Trans. A* **16A**, 249 (1985)
- S. Suresh, R.O. Ritchie, A geometric model for fatigue crack closure induced by fracture surface morphology. *Metall. Trans. A* **13A**, 1627 (1982)
- S. Suresh, R.O. Ritchie, Some consideration on the modelling of oxide-induced crack closure using solution for a rigid wedge inside a linear elastic crack. *Scripta Metall.* **17**, 575 (1983)
- J. Weertman, Fatigue Crack Growth in Ductile Metals. *Mechanics of Fatigue*, AMD-Vol. **47**, (ed.) by T. Mura, ASME (1982), p. 11

## Growth Characteristics of Small Fatigue Cracks

KWAI S. CHAN

Mechanical Engineering Division, Southwest Research Institute®, San Antonio, TX, USA

## Synonyms

Mechanically small cracks; Microcracks; Microstructurally small cracks; Physically small cracks; Short cracks

## Definition

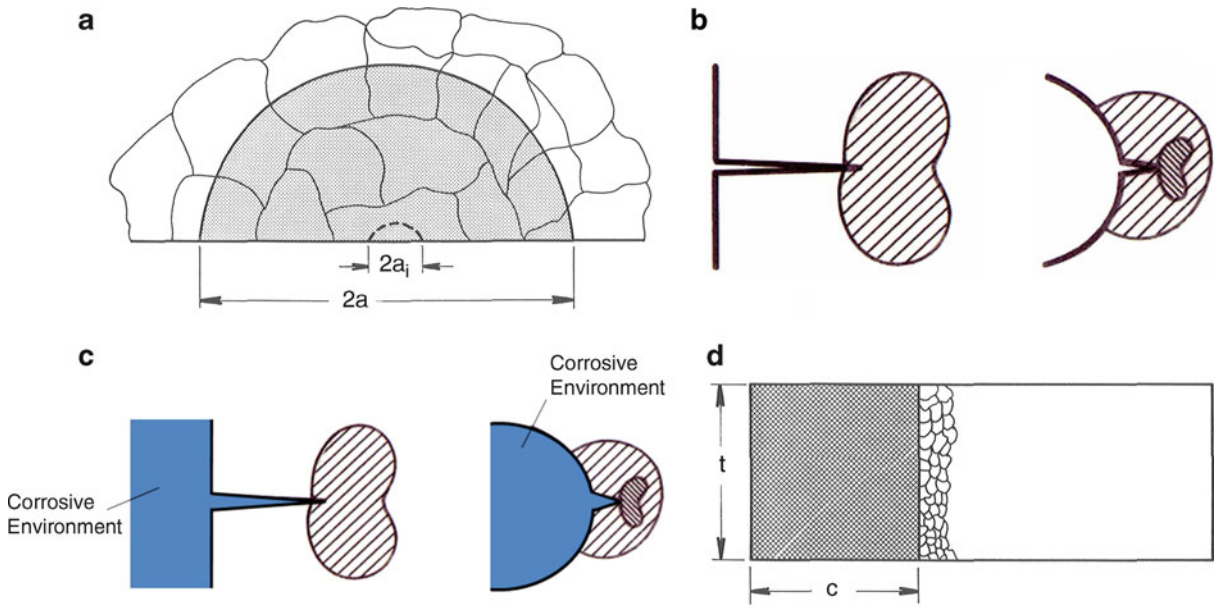
Fatigue cracks whose dimensions are small compared to the underlying microstructural unit size, often referred to

as “small” cracks, exhibit growth behavior that is significantly different from what would be expected based on conventional (i.e., large-crack) fatigue crack growth (FCG) data or analysis techniques (► [Growth Characteristics of Large Fatigue Cracks](#), ► [Fatigue](#)). In particular, small fatigue cracks often grow faster than corresponding large cracks at the same nominal value of the stress intensity factor range,  $\Delta K$ , or grow at  $\Delta K$  levels below the threshold value,  $\Delta K_{th}$ , determined from traditional large-crack test methods (Davidson et al. 2003). Therefore, a structural life assessment based on large-crack FCG analysis methods can be non-conservative if the initial crack size is sufficiently small that the life is dominated by small-crack growth.

Several types of small cracks exist because different mechanisms are responsible for their different growth behaviors. A small crack is microstructurally small when all crack dimensions are small in comparison to the characteristic microstructural dimensions. The relevant microstructural feature that defines this scaling may vary from material to material, but the most common microstructural scale is the grain size, as illustrated schematically in [Fig. 1a](#) (Chan and Lankford 1988). The small fatigue crack and its crack-tip plastic zone may be embedded completely within a single grain, or the crack size may be on the order of a few grain diameters. The terms *small crack* and *short crack* both appear in the literature, and sometimes the two appear to be used interchangeably. However, the two terms have distinct meanings among many researchers. In the US research community, the currently accepted definition for a “small” crack requires that all physical dimensions (in particular, both the length and depth of a surface crack) are small in comparison to the relevant length scale. The relevant length scale, and hence the specific physical dimensions, vary with the particular material, geometry, and loading of interest. In contrast, a crack is defined as being “short” when only one physical dimension (typically, the length of a through-crack) is small in comparison to the length scale. Because the crack front of a short crack interrogates many different grains, it is not usually subject to strong microstructural effects. However, it should be noted that this nomenclature distinction has not always been observed in the literature, and some current authors (especially in Europe) employ the terms with nearly reverse meanings (Davidson et al. 2003).

A crack is generally considered to be “mechanically small” when all crack dimensions are small compared to characteristic mechanical dimensions. The relevant mechanical feature is typically a zone of plastic deformation, such as the crack tip plastic zone or a region of local





**Growth Characteristics of Small Fatigue Cracks, Fig. 1** Schematics of the different local crack-tip environment encountered by a fatigue crack: (a) a microstructurally small crack (Chan and Lankford 1988), (b) mechanically small cracks (Davidson et al. 2003), (c) chemically small cracks, and (d) a short or large crack (Chan and Lankford 1988)

plasticity at some mechanical discontinuity (e.g., a notch). The crack may be fully embedded in the plastic zone, or the plastic zone size may simply be a large fraction of the crack size. Both types of mechanically small cracks are shown in Fig. 1b. The regimes of mechanically and microstructurally small cracks can overlap (Davidson et al. 2003). Many microstructurally small cracks are also mechanically small, although a microstructurally small crack in a large grain or single crystal material may be mechanically small. Many mechanically small cracks of practical importance are microstructurally large.

Chemically small cracks exhibit higher fatigue crack growth rates compared to large cracks at equivalent nominal environment and mechanical loading conditions, as shown in Fig. 1c. The chemical size effect has been reported for 4130 steel in an aqueous NaCl environment (Gangloff 1985). Corrosion-fatigue crack growth rates from small surface cracks (0.1–1 mm deep), as well as short through-thickness edge cracks (0.1–3 mm), are appreciably faster than corrosion-fatigue crack growth rates from large through-thickness cracks (25–40 mm) in standard compact tension specimens. Moreover, all of the corrosion-fatigue crack growth rates in NaCl are enhanced compared to small and large fatigue cracks in a moist laboratory air environment. The chemical small-crack effect can occur over a much larger range of crack sizes (up to 5 mm) than fatigue small cracks.

**Growth Characteristics of Small Fatigue Cracks, Table 1**  
Small fatigue crack classification (Davidson et al. 2003)

Description	Approximate Dimension	Reason for small-crack behavior
Mechanically small	$\frac{a}{r_p} < 4 - 10$	Excessive plasticity R-ratio shift
Microstructurally small	$\frac{a}{D} < 2 - 10$	Crack closure Excessive crack tip plasticity
Chemically small	$a < 5 \text{ mm}^a$	Liquid environment
Small crack size approximation	$a < a_o$	Derived from Kitagawa diagram

$2a$  = Surface crack length

$r_p$  = Plastic zone size at crack tip or notch plastic field size

$D$  = A relevant microstructural dimension

<sup>a</sup>Critical size is a function of frequency and reaction kinetics

For comparison, Fig. 1d shows a short or a large fatigue crack, which crack tip encounters a large number of grains in the crack-tip plastic zone (Chan and Lankford 1988). Because the terms used to define small cracks overlap and may be confusing, some rules of thumb are given in Table 1 to clarify the descriptions of size effects as well as help in approximating the conditions under which cracks may be considered as small.

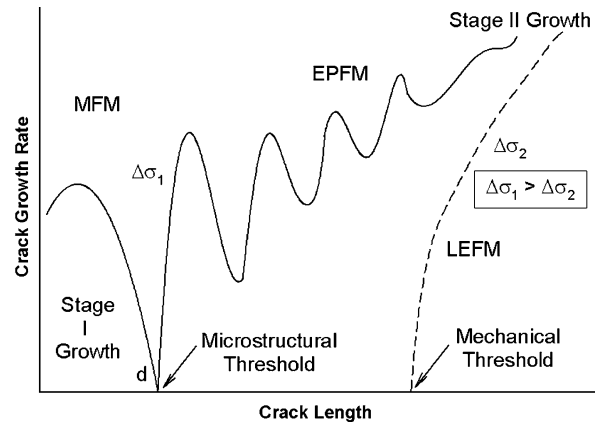


## Scientific Fundamentals

Small crack represents the bridge between fatigue crack nucleation and large fatigue crack growth. Crack initiation, which sets the limit on the minimum size of a small fatigue crack, is a complex process that can lead to a crack that is extremely small ( $\approx 1 \mu\text{m}$ ) or one that appears to “pop in” and rapidly grows to lengths up to  $\approx 100 \mu\text{m}$  long. Planar slip materials may concentrate slip and form persistent slip bands that eventually become cracks. Coarse-grained materials are more prone to slip band cracks than fine-grained materials. In some alloys, the apparent grain size may not be an accurate view of the effective grain size characterizing the slip distance. An example of this phenomenon is a ferritic-pearlitic steel, where the underlying alpha grain has only one crystallographic orientation (Davidson et al. 2003), or a grain existing in the gamma phase before a martensitic transformation, which is larger than any dimension in the transformed structure. Texture may enhance fatigue crack initiation if groups of grains are oriented for easy slip within the applied stress field. For example, prior  $\beta$  grains in  $\alpha + \beta$  Ti alloys transform to groups of similarly oriented  $\alpha + \beta$  lamellar colonies that exhibit similar slip behavior and easy slip transmission across colony boundaries. The effective slip distance of the similarly oriented  $\alpha + \beta$  lamellar colonies is the size of the prior  $\beta$  grains and not the smaller lamellar colony size. A group of similarly oriented grains that exhibit cooperative slip behavior over a distance larger than the characteristic distance of individual grains is often referred to as super-grain, whose larger slip distance often leads to easy crack initiation and lower fatigue life.

The nature of the slip characteristics often dictates the fracture surface topography near the origin of the crack. Stage I fatigue cracks are initiated by the intrusion-extrusion mechanism from persistent slip bands, and growth at an angle to the stress axis. The interaction of Stage I crack with microstructure may require treatment by microstructural fracture mechanics (MFM), as illustrated in Fig. 2 (Miller 1993). As the Stage I cracks grow larger, they alter the crack path to grow more perpendicular to the stress axis, eventually becoming Stage II fatigue cracks.

Fatigue cracks in commercial materials start at metallurgical stress concentrations such as inclusions and pores. Surface finish has a strong influence on fatigue crack initiation and residual stresses due to surface preparation can have considerable influence on the number of cycles to initiation. The anisotropy and texture of a material also have strong influences. Cracks can initiate also in grain boundaries, with or without the influence of environment,

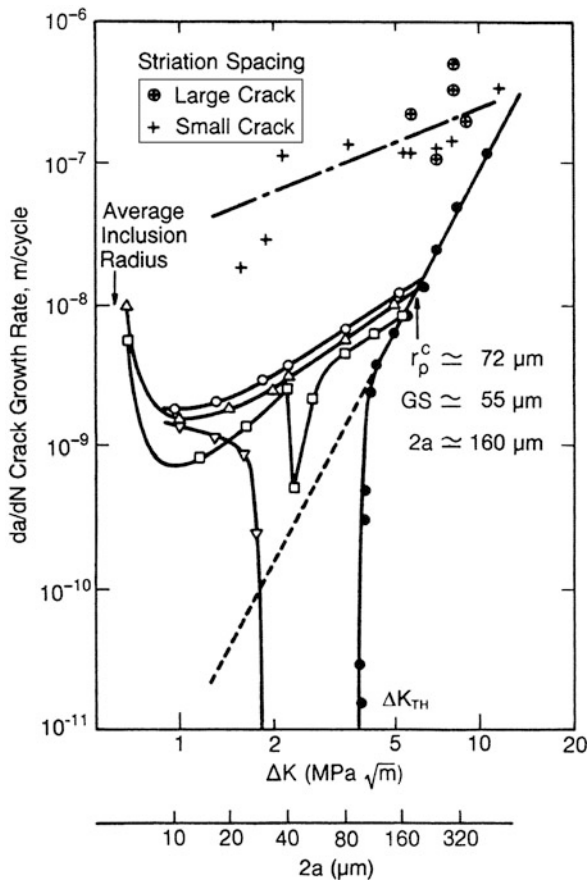


**Growth Characteristics of Small Fatigue Cracks, Fig. 2** Crack propagation rate versus crack length for microstructurally small to mechanically small cracks (Miller 1993)

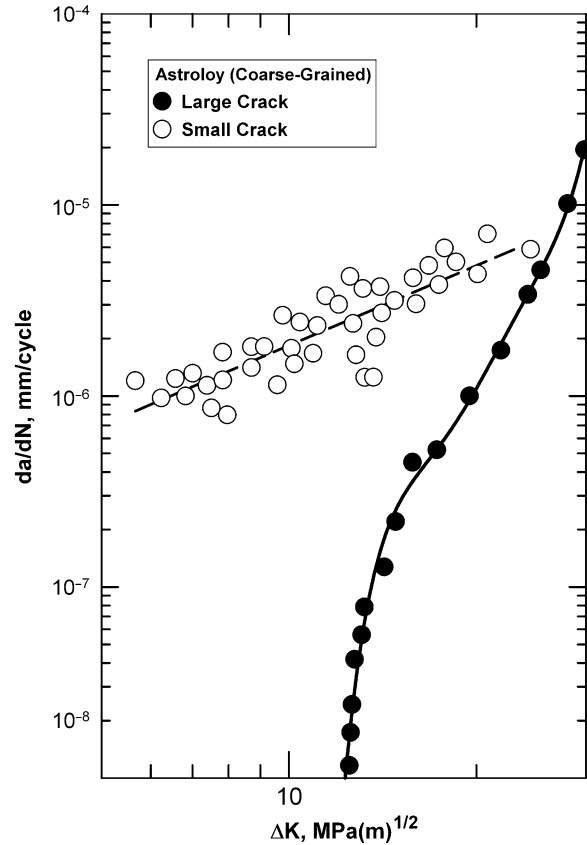
or brittle intermetallics situated along grain boundaries, or due to the chemical segregation of impurities to the grain boundary.

For initiation at defects, the fatigue cracks formed and propagated as Stage II cracks normal to the principal stress direction. Typical crack growth data for microstructurally small cracks initiated at particles and propagate as Stage II cracks at stresses above the fatigue limit are shown in Fig. 3 for a 7075 aluminum alloy, along with traditional large-crack data for the same alloy (Lankford 1982). Figure 3 shows that small-crack growth occurs at nominal  $\Delta K$  values below the large-crack threshold. Furthermore, small-crack growth rates are often higher than that would be predicted by extrapolating the large-crack Paris equation and the apparent linear slope to lower  $\Delta K$  values. Despite the higher growth rates, crack arrest, temporary or otherwise, can occur at microstructural barriers such as grain or phase boundaries when the crack size is on the order of the microstructural unit size. However, not all small cracks arrest or even slow down at the microstructural barriers. As the small cracks continue to grow, the crack growth rate data often merge with the large-crack data, indicating small cracks eventually become large cracks with increasing crack lengths.

Because of large-scale crack-tip plasticity, elastic-plastic fracture mechanics may be required to treat Stage II growth of mechanically small fatigue cracks. In contrast, the growth behaviors of large cracks are typically treated by linear-elastic fracture mechanics (LEFM), as shown in Fig. 2 (Miller 1993). Stage I cracks are fairly common in  $\alpha + \beta$  Ti alloys and Ni-based superalloys as both exhibit planar slip at ambient and intermediate temperatures.



**Growth Characteristics of Small Fatigue Cracks, Fig. 3**  
Typical FCG data for microstructurally small cracks and large cracks (Lankford 1982)



**Growth Characteristics of Small Fatigue Cracks, Fig. 4**  
Fatigue crack growth rate data of small and large cracks for a coarse-grained Ni-based alloy (Astroloy) in the near-threshold region of the large crack (Modified from Lankford and Hudak 1987)

Figure 4 shows the crack growth response of Stage I small cracks in a coarse-grained Ni-based superalloy (Astroloy). These small cracks grow at higher rates at  $\Delta K$  levels below the large crack threshold. In general, highly planar slip materials such as Ni-based superalloys and  $\alpha + \beta$  Ti alloys tend to exhibit very rough fracture surfaces near the crack origins, becoming smoother as the crack extends and propagate as Stage II cracks. In contrast, multiple slip materials generally evidence less of this transition as Stage II fatigue cracks are more prevalent in these materials.

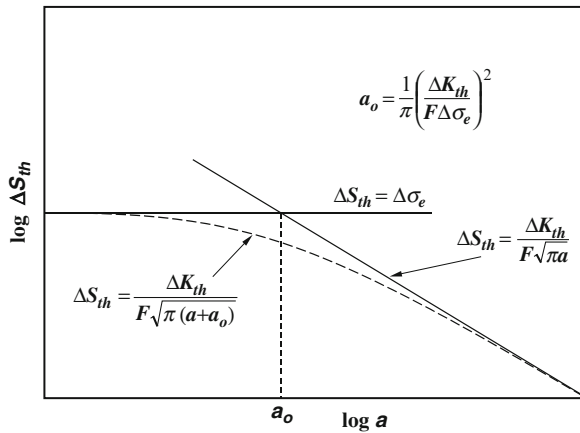
The relationship between small cracks, large cracks, and crack initiation is illustrated by the Kitagawa diagram (Fig. 5) (Kitagawa and Takahashi 1967). This diagram integrates the traditional large crack fracture mechanics threshold,  $\Delta K_{th}$ , with the traditional smooth specimen endurance limit,  $\Delta \sigma_e$  (► [Fatigue Limit](#)). The region on the Kitagawa diagram both above the horizontal

endurance limit line and above the sloping threshold line is “unsafe” in that fatigue failure is predicted to occur. The region below both lines is “safe” in that fatigue failure is predicted not to occur. The intersection of the two lines is commonly denoted as  $a_0$ , which is given by El Haddad et al. (1979)

$$a_0 = \frac{1}{\pi} \left[ \frac{\Delta K_{th}}{F \Delta \sigma_e} \right]^2 \quad (1)$$

where  $\Delta K_{th}$  is the large-crack FCG threshold,  $\Delta \sigma_e$  is the fatigue endurance limit, and  $F$  is the boundary-correction factor (► [Stress Intensity Factors](#)).

The Kitagawa diagram demonstrates that cracks smaller than  $a_0$  must be able to grow at nominal stress intensity factors that are less than the large crack threshold, since smooth specimens fail by the initiation and



**Growth Characteristics of Small Fatigue Cracks, Fig. 5**

Kitagawa diagram depicting regions of crack initiation and crack growth in a double logarithm plot of stress range versus crack length (Kitagawa and Takahashi 1967; El Haddad et al. 1979)

growth of microcracks to failure at stresses greater than the endurance limit (► [Fatigue Limit](#)). On the other hand, El Haddad and others (El Haddad et al. 1979) have proposed that the safe/unsafe boundary is properly represented not by the two intersecting straight lines, but by the curving dashed line in Fig. 5 that provides a smooth transition between the endurance limit and threshold lines. This construction implies that cracks smaller or slightly larger than  $a_0$  can also grow both below the large crack threshold and below the endurance limit stress.

The small crack growth response at  $\Delta K$  levels below the large crack thresholds can be attributed to a lack of similitude in the crack driving force, microstructure, or both.

## Crack Driving Force

The assumption of small-scale yielding in linear-elastic fracture mechanics (LEFM) is violated in small cracks under several fatigue conditions, which include fatigue loading under high stresses in relation to the yield strength, and for crack propagation within a local plastic strain field, such as that of a notch, pore, particle, or an optimally orientated surface grain (Davidson et al. 2003). Under these circumstances, the driving force for fatigue crack growth is not accurately treated by LEFM and the large-scale yielding experienced at the crack tip must be used to correct the crack driving force. Elastic-plastic fracture mechanics (EPFM) has been used to address this breakdown of mechanical similitude. When the crack size is on the order of the microstructural unit or grain size,

microstructural fracture mechanics (MFM) is required to address this breakdown of similitude. Figure 2 illustrates the crack size regimes where MFM, EPFM, and LEFM are applicable for fatigue cracks (Miller 1993).

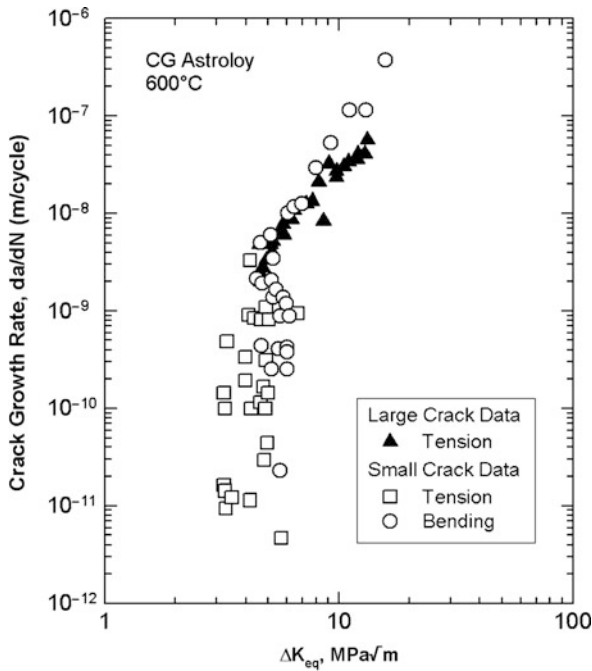
Several relations have been developed to compute the local crack driving force for short and small cracks subjected to large-scale yielding (Davidson et al. 2003). These relations are formulated either expressly or implicitly on the basis of  $\Delta J$ , the cyclic J integral. The  $\Delta J$ -integral was first applied to elastic-plastic fatigue under gross yielding conditions. Subsequent work has demonstrated that the  $\Delta J$  approach is applicable to fatigue loading where gross yielding is absent but the stresses are high in relation to the yield strength of the material. A non-linear crack tip parameter describing both short and small cracks is the equivalent stress intensity factor range,  $\Delta K_{eq}$ , which is defined in terms of the  $\Delta J$  integral as (Davidson et al. 2003)

$$\Delta K_{eq} = \left[ \frac{E \sigma_y \delta}{1 - \eta^2} \right]^{1/2} \quad (2)$$

where  $E$  is Young's modulus,  $\sigma_y$  is the yield stress, and  $\delta$  is the crack-tip opening displacement (CTOD);  $\eta = 0$  for plane stress and  $\eta = \nu$  plane strain. In a series of studies, the crack-tip opening displacements of small fatigue cracks were measured. The  $\Delta K_{eq}$  parameter computed based on the experimental crack-tip opening displacement via (2) was successful in collapsing small and large crack data, as shown in Fig. 6 for Astroloy. In addition, the validity of (2) for short and small cracks has been demonstrated for other structural alloys (Davidson et al. 2003). Instead of using experimental values of CTOD, (2) can be used in conjunction with theoretical values that are computed based on the analytical or numerical methods.

Several crack closure mechanisms have been identified in metals and structural alloys, which include contact of the crack surfaces resulting from crack-wake plasticity, oxide wedging, and fracture-surface asperities (Suresh and Ritchie 1984). The plastic-wake phenomenon has been modeled using the continuum mechanic approach. Crack closure by fracture surface asperities is often referred as roughness-induced closure (Suresh and Ritchie 1984). Both oxide wedge and roughness-induced crack closure are microstructure-sensitive; as a result, they have been modeled at the microstructural scale levels via the micromechanical approach.

Plasticity-induced crack closure in short cracks was modeled by Newman, who used the Dugdale model to represent the crack-tip plastic zone and compute the residual plastic strain on the surfaces of the crack wake (Newman 1995). Contact of the fracture surfaces during



**Growth Characteristics of Small Fatigue Cracks, Fig. 6**  
Correlation of small-crack and large-crack data by modifying the driving force parameter,  $\Delta K_{eq}$ , to account for local crack-tip plasticity on the basis of CTOD or  $\Delta J$  (Davidson et al. 2003)

loading leads to delayed crack opening, thereby reducing the stress intensity range between opening and maximum loads. The amount of plasticity-induced crack closure depends on the stress state at the crack tip, the crack-tip strain, the amount of plastic strain in the crack wake, and the residual stress in the plastic zone. In addition to plasticity-induced crack closure, other closure mechanisms are important in influencing the near-threshold behavior of large cracks. These include oxide-induced crack closure, roughness-induced crack closure, and crack closure resulting from crack branching and deflection (Suresh and Ritchie 1984). These closure processes are often referred to as crack-tip shielding mechanisms that reduce the local driving force for crack growth by decreasing the local  $\Delta K$  at the crack tip during the fatigue cycle. In general, the closure mechanisms explain the small-crack effects on the basis that crack closure is absent or small at small crack length but crack closure increases in magnitude as a crack gets longer and develops a plastic wake with or without the presence of other shielding mechanisms to reduce the near-tip stress intensity factor, leading to crack-size and mean stress dependence of FCG rates in the small crack region.

## Fatigue Crack Growth Resistance

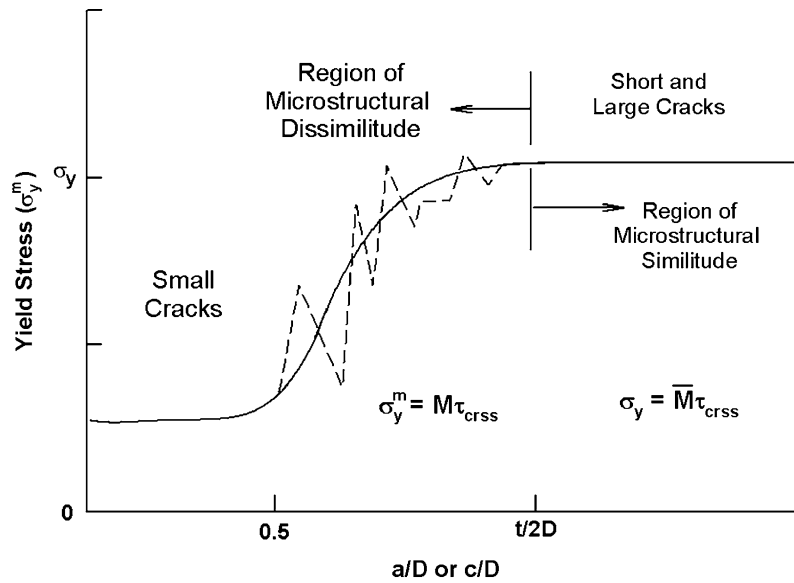
Instead of a size-dependent crack driving force, small crack effects can be considered on the basis of a size-dependent fatigue crack growth resistance due to changes in the crack tip microstructural environment with crack length. For example, small surface cracks are more susceptible to microstructural effects than through-thickness large cracks because of the different number of units encountered at their crack tips (Chan and Lankford 1988). As illustrated in Fig. 1a, the number of grains encountered by a small crack of initial length,  $2a_i$ , embedded within a single grain is unity. However, as the half-length,  $a$ , of the small crack increases, the number of grains along the crack front increases proportionately, and ultimately approaches that of a through-thickness large crack, shown in Fig. 1d. Because of this sampling effect, the mechanical properties (e.g., yield stress, number of slip planes, etc.) of the crack tip plastic zone of a small crack varies with crack length and approaches that of a large crack (average properties) as the crack length increases.

The variation of yield stress of the crack tip plastic zone with crack size for small cracks is illustrated in Fig. 7 (Chan and Lankford 1988). Fatigue cracks usually initiate in an optimally oriented surface grain with a low Taylor factor (e.g.,  $M = 2$ ), where  $M$  is the Taylor factor. As the crack length increases, the value  $M$ , and therefore the yield stress, increases due to increasing numbers of grains encountered by the crack tip. The ratio of the plastic zone size to crack size is on the order of unity when the small crack is embedded within a grain; thus, the small crack is subjected to a large-scale plastic field. As the crack length increases, the plastic zone size to crack size decreases due to increasing yield stress and it approaches that of a large crack.

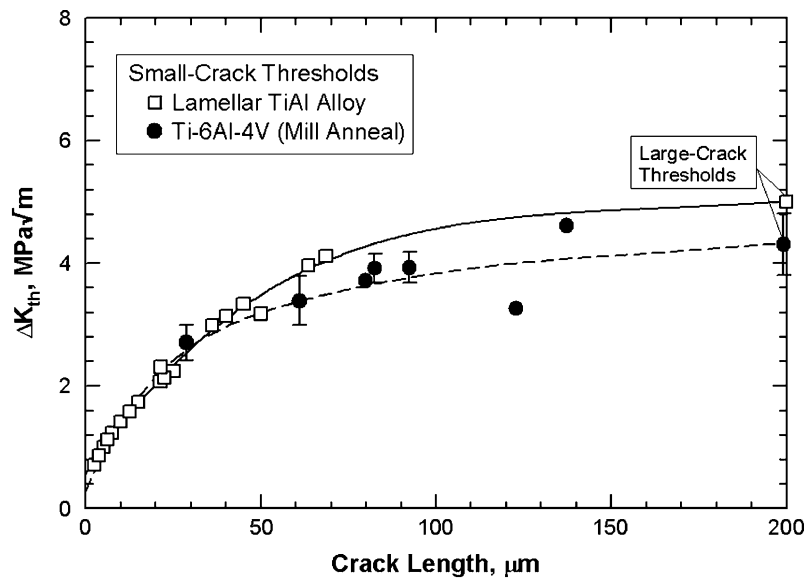
Furthermore, the extent of crack closure and shielding mechanisms as measured by the crack closure load increases with increasing crack length or depth as a fatigue cracks grows, which translates into an increase in the apparent fatigue crack growth thresholds. As shown in Fig. 8, extensive experimental evidence indicates that the fatigue crack threshold,  $\Delta K_{th}^{SC}$ , for small cracks increases with increasing crack length and approaches the large-crack threshold,  $\Delta K_{th}^{LC}$ . This threshold dependence on crack size can be presented in terms of  $a_o$  as follows:

$$\Delta K_{th}^{SC}(a) = \Delta K_{th}^{LC} \sqrt{\frac{a}{a + a_o}} \quad (3)$$

which correlates well with both the threshold data of various crack sizes for several alloys. For crack depth less



**Growth Characteristics of Small Fatigue Cracks, Fig. 7** A schematic showing the dependence of yield stress within the plastic zone on the ratio of the crack size ( $a$ ) to grain size ( $D$ ) for small, short, and large cracks (Chan and Lankford 1988)

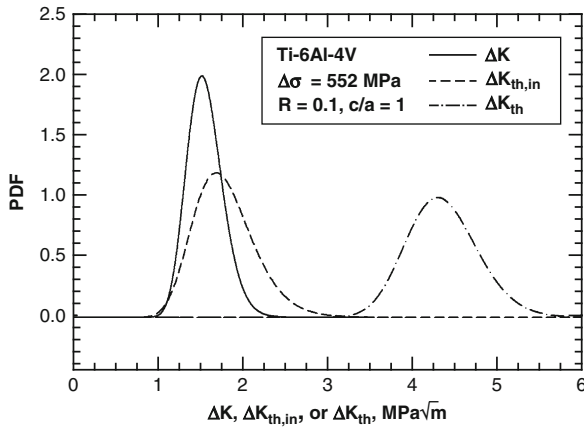


**Growth Characteristics of Small Fatigue Cracks, Fig. 8** Variation of the fatigue crack growth threshold of small crack as a function of crack size (Davidson et al. 2003)

than half of a grain size, the small-crack FCG threshold corresponds to the intrinsic FCG threshold. The value of  $\Delta K_{th}^{sc}$  increases with crack depth or length and approach to the large-crack FCG threshold,  $\Delta K_{th}^{lc}$ , when a becomes large. Figure 8 illustrates the increase of the apparent

FCG threshold of small cracks observed in lamellar  $\alpha_2 + \gamma$  TiAl alloys and mill-annealed Ti-6Al-4V.

Because of the variability of crack growth rates of small cracks, the FCG threshold can appear to be absent for the grain-sized small cracks because the stress intensity range



**Growth Characteristics of Small Fatigue Cracks, Fig. 9**

Probability distribution functions of intrinsic threshold, large-crack threshold, and the stress intensity range of grain-facet-sized small cracks for Ti-6Al-4V with a duplex microstructure at 25°C (Chan and Enright 2006)

due to the applied loads always exceeds the intrinsic threshold and the large-crack threshold for certain microstructures or grain size distributions. One such example is for Ti-6Al-4V with a duplex microstructure. The probability distribution function (PDF) for the applied stress range acting on a grain-sized crack after initiation is shown as the solid curve in Fig. 9 for  $\Delta\sigma = 552$  MPa. For comparison, Fig. 9 also presents the PDF for the intrinsic threshold and the PDF for the large-crack threshold. An analysis by Chan and Enright (2006) indicated that the applied  $\Delta K$  is well below the large-crack threshold that the probability of  $\Delta K < \Delta K_{th}$  is essentially one ( $>0.999999$ ) for stress levels at  $\Delta\sigma = 552$  MPa and 747 MPa. Thus, the probability of propagation for the grain-sized cracks would be essentially zero (less than  $1E-6$ ) if the large-crack FCG threshold controls the growth behavior. In contrast, the PDF for the applied  $\Delta K$  overlaps that for the intrinsic threshold,  $\Delta K_{th,in}$ . The probability of  $\Delta K > \Delta K_{th,in}$  is 0.307 at  $\Delta\sigma = 552$  MPa, but is increased to 0.768 at  $\Delta\sigma = 747$  MPa. The results indicate that only 30.7% of the grain-sized cracks initiated at  $\Delta\sigma = 552$  MPa would propagate while the remaining 69.2% of the cracks would arrest. By the same token, 76.8% of the grain-sized cracks initiated at  $\Delta\sigma = 747$  MPa would propagate, while the remaining 23.2% would arrest.

The continual growth of a small crack ( $\Delta K > \Delta K_{th,in}$ ) at  $\Delta\sigma = 552$  MPa is illustrated in Fig. 10, which compares the  $\Delta K$  and  $\Delta K_{th}^{SC}$  of the small crack as a function of crack extension. The value of  $\Delta K_{th}^{SC}$  was computed via (1) and (3) and the  $\Delta K_{th}^{LC}$  value was taken to be evaluated

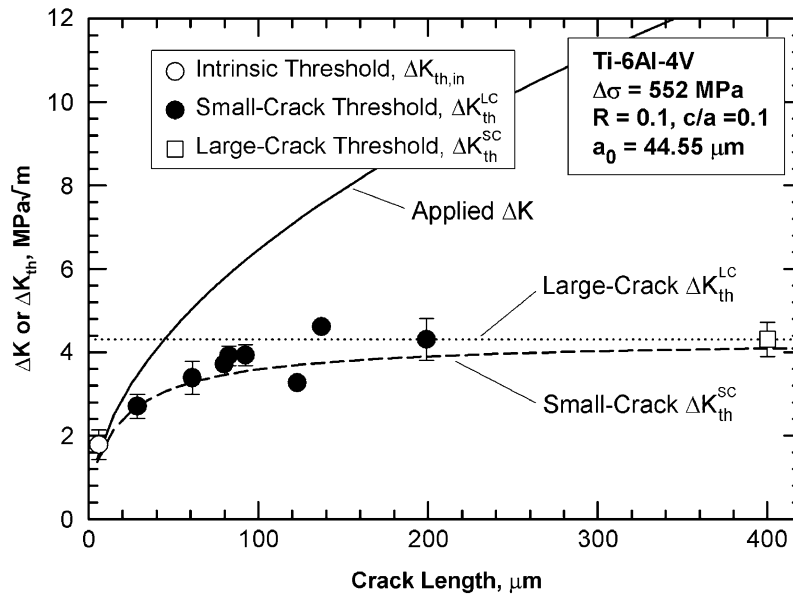
$a_0 = 45$   $\mu\text{m}$ . The applied  $\Delta K$  of the small crack increases with increasing crack lengths according to the solid curve shown in Fig. 10. As the crack length increases, the FCG threshold of the small crack also increases as depicted by dashed curve (Fig. 10) and approaches to the large-crack threshold at large crack length ( $>400$   $\mu\text{m}$ ). As indicated in Fig. 10, the predicted  $\Delta K_{th}$  values are in agreement with experimental data for both the small-crack and large-crack regime. For crack lengths greater than 45  $\mu\text{m}$ , the applied  $\Delta K$  exceeds the intrinsic and the large-crack FCG thresholds. Such a small crack would appear to propagate to failure without a threshold since it propagates at  $\Delta K$  levels below the large-crack FCG threshold. It is noted that 30.7% of the grain-sized cracks in Ti-6Al-4V with the duplex microstructure behave in this manner for  $\Delta\sigma = 552$  MPa. The number of these propagating small cracks is expected to decrease as the stress level approaches the fatigue limit. Whether or not a grain-sized crack will propagate after initiation has been examined by computing the probability of the  $\Delta K < \Delta K_{th,in}$ ,  $\Pr(\Delta K < \Delta K_{th,in})$ , on the basis of the grain size distribution and by varying the stress range.

## Key Applications

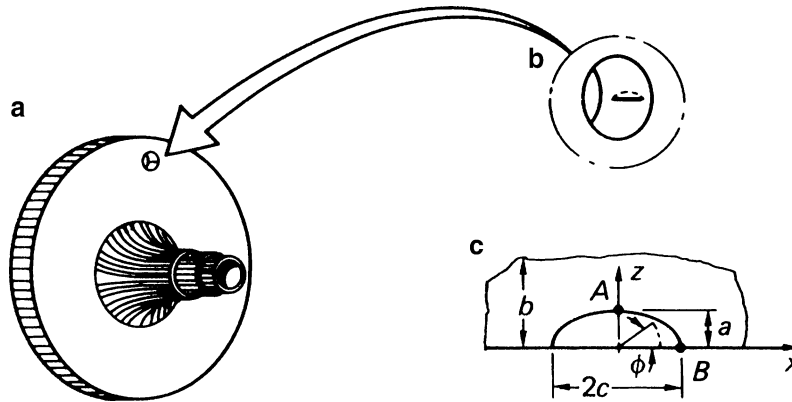
Small crack behavior will not be an important issue for applications in which initial defects are large or fatigue cracks are large at initiation. Small cracks are not a consideration for many traditional mechanical engineering design/analysis applications employing damage tolerance concepts (such as aircraft structures), because the initial flaw size, which may be based on conventional nondestructive detection evaluation (NDE) inspection limits, is usually much larger than the small-crack regime. Furthermore, small crack behavior will not be practically significant for some high cycle fatigue (HCF) problems in which a very large fraction (perhaps  $>90\%$ ) of the total life is consumed by the formation of even a microscopic crack, which then grows quickly to failure under the high frequency loading.

However, small crack behavior must be considered in numerous other fracture mechanics applications. When damage tolerance methods are applied to more highly stressed safety-critical structures, tolerable flaw sizes are generally smaller and NDE requirements are stricter. These applications have historically been treated with safe-life methods based on bulk damage strain-life or stress-life analyses. However, the total life in many strain-life applications is often dominated by the growth of small cracks, especially in the low-cycle fatigue (LCF) regime where crack formation occurs very early in life and final crack sizes are still relatively small. Therefore, the





**Growth Characteristics of Small Fatigue Cracks, Fig. 10** Increase of the applied  $\Delta K$  with crack length is faster than that of the small-crack threshold,  $\Delta K_{th}^{SC}$ .  $\Delta K_{th}^{SC}$  increases from a value of  $\Delta K_{th,in}$  (1.75 MPa√m) to the large-crack threshold  $\Delta K_{th}^{LC}$  (4.2 MPa√m) as crack length increases (Chan and Enright 2006)

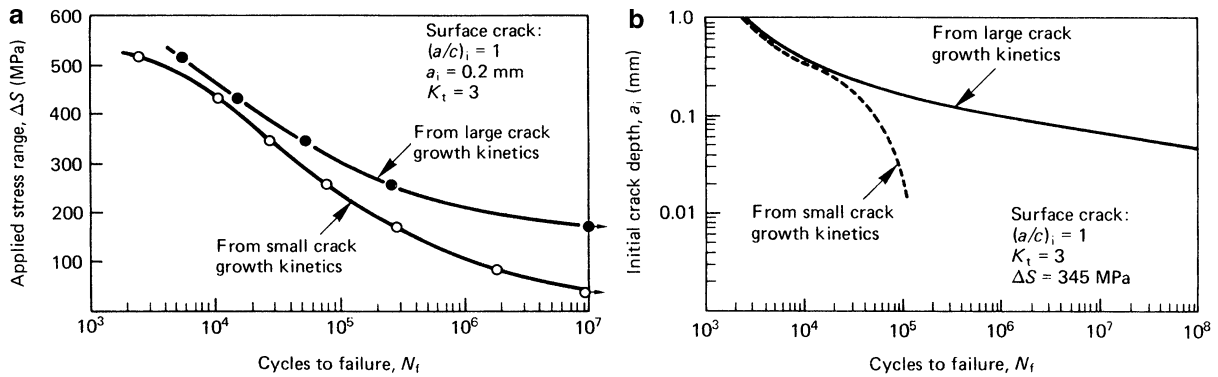


**Growth Characteristics of Small Fatigue Cracks, Fig. 11** Schematics of a semi-circular surface flaw at the cooling hole in the web of a Ni-based alloy (Astroloy) disk: (a) disk, (b) crack location, and (c) semi-circular crack of surface length  $2c$  and crack depth  $a$  (Lankford and Hudak 1987)

damage accumulation process during LCF, which is often treated as an “initiation” problem, is often actually a small-crack growth process. Small-crack analysis techniques may provide valuable new insights into some difficult LCF lifing problems, such as load interaction effects.

Durability assessments of aircraft structures based on fracture mechanics typically employ much smaller initial flaw sizes than for standard damage tolerance assessments [e.g., 0.005 in. (0.127 mm) rather than 0.050 in.

(1.27 mm)]. The initial flaw size for a durability assessment may also be based on an equivalent initial flaw size that is back-calculated from total life. In either case, it is possible that the initial flaw may exhibit small-crack growth characteristics. The application of small-crack damage approach to LCF lifing of an engine disk was demonstrated by Lankford and Hudak (1987), for a semi-circular surface flaw at a cooling hole in the web of a Ni-alloy disk (Fig. 11) using small-crack and



**Growth Characteristics of Small Fatigue Cracks, Fig. 12** Predicted fatigue life responses of small and large cracks for the Ni alloy disk: (a) stress range versus life, and (b) initial crack depth versus life (Lankford and Hudak 1987)

large-crack data (Fig. 4) for coarse-grained Astroloy and an elastic stress concentration factor of 3 ( $K_t = 3$ ). A crack growth equation that includes all three stages (I, II, and III) of FCG was used to fit to the large-crack data and the same equation was modified by dropping the Stage I term and fitted to the small-crack data. The predicted fatigue lives as a function of applied stress range are shown in Fig. 12a. In the short life and high stress range region, the small-crack and large-crack curves exhibit only small difference. As the stress range decreases, the difference between the small-crack curve and the large-crack curve becomes increasingly large due to the growth kinetics and the absence of a threshold for the small cracks. The influence of the initial crack size on the predicted FCG lives for small and large cracks is illustrated in Fig. 12b, which shows significantly lower lives for the small cracks at decreasing initial crack sizes. In contrast, both the small-crack and large crack model predict essentially identical LCF lives when the initial crack size is relatively large ( $>0.3$  mm). The predicted lower fatigue lives for small-crack dominated LCF failure have been reported for a number of Ti-based and Ni-based alloys (Jha et al. 2007).

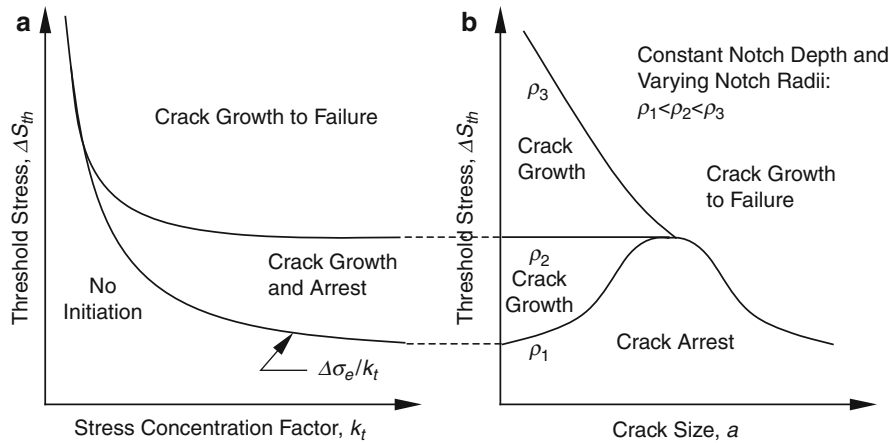
Small crack behavior may also be very important for HCF applications when various damage states (e.g., foreign object damage, fretting, or superimposed LCF) are also operative (Hudak et al. 2002). Although pure HCF (constant amplitude loading with no superimposed damage) may induce cracking relatively late in life, these damage states can produce a small crack relatively early in life. The original damage source itself may not be able to grow the crack to failure, but the critical question is whether the crack will propagate as a result of the HCF loading. Subsequent crack growth will probably depend on the magnitude of  $\Delta K_{th}(a)$  because any crack growth will quickly

result in failure due to the high frequency of loading; thus, the potential crack size effect on threshold may be very significant.

Foreign object damage (FOD) is capable of inducing relatively sharp notches that exhibit a wide range of depths (0.1–10 mm) and radii (0.1–2.0 mm). Classical fatigue research has shown that sharp notches are less damaging than one might anticipate based on their high theoretical elastic stress concentration factor because sharp notches tend to initiate microcracks that eventually arrest as they grow out of the influence of the highly concentrated notch stress field. Recent understanding of the unique behavior of small fatigue cracks has provided a means of explaining crack arrest at notches, and initial success using fracture mechanics to estimate the propagation and arrest of small cracks at notches in HCF design.

### Worst-Case Notch

A treatment of small-crack effects and the associated conditions for non-propagation of cracks in the stress gradients of notches can lead to the definition of a limiting “worst-case notch.” Such an approach has been applied to the problem of variable weld-toe notches in offshore structures. It could also provide a viable engineering solution to the problem of dealing with the large variation in FOD-induced notches that is encountered in turbine engine blades (Hudak et al. 2002). The “worst-case notch” concept is illustrated in Fig. 13 where the nominally applied threshold stress for a notched member is shown as a function of both the elastic stress concentration factor  $k_t$  and crack size “a.” As can be seen in Fig. 13a, the threshold stress, computed by dividing the endurance limit ( $\Delta\sigma_e$ ) by  $k_t$ , continually decreases with increasing  $k_t$  (and notch severity). However, worst-case notch theory



**Growth Characteristics of Small Fatigue Cracks, Fig. 13** Worst-case notch concept showing: (a) limiting threshold stress with increasing notch stress concentration, and (b) threshold stress versus crack size and notch severity characterized in terms of notch root radii,  $\rho_1 < \rho_2 < \rho_3$ , for a constant notch depth (Hudak et al. 2002)

predicts a limiting  $k_t$ , termed  $k_{w}$ , above which further increases in  $k_t$  no longer result in decreases in the threshold stress due to the occurrence of crack arrest and non-propagating cracks. These curves define boundaries separating three distinct regimes where: (1) crack initiation will not occur; (2) crack initiation will occur, followed by crack arrest; and (3) crack initiation followed by crack growth will result in failure of the notched member. These same regimes are shown as a function of crack size in Fig. 13b, for a fixed notch depth and three different notch root radii,  $\rho_1$ ,  $\rho_2$ ,  $\rho_3$ . As indicated in Fig. 13b, the threshold stress increases as the radius increases and there exists a limiting radius,  $\rho_w$ , above which crack arrest cannot occur. In general, the threshold stress is a function of the FOD-notch geometry, loading mode, and material variables as follows:

$$\Delta S_{th} = \Delta S_{th}(k_t[b, \rho], a, a_0, \lambda) \quad (5)$$

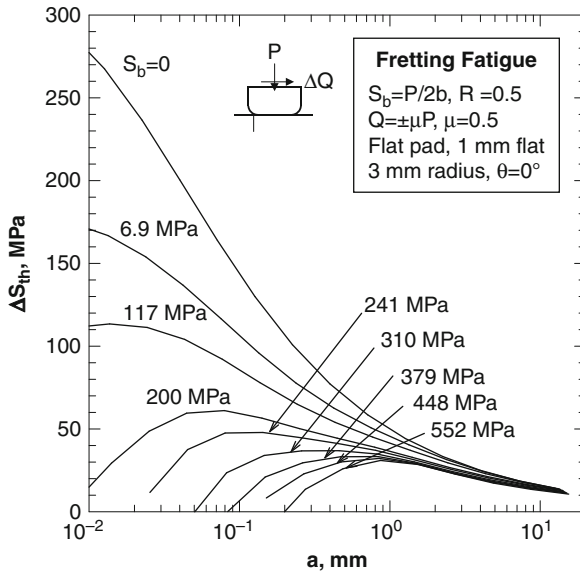
where  $\lambda$  is a parameter that depends on loading mode (e.g., tension vs. bending) and  $a_0$  is the small crack parameter (El Haddad et al. 1979). Equation (5) can be solved using the input and methods discussed in detail elsewhere (Hudak et al. 2002).

### Worst-Case Fret

Fretting fatigue occurs when contacting surfaces experience small amplitude cyclic displacements while simultaneously being subjected to bulk cyclic stresses. The small amplitude surface displacements generate high cyclic shear stresses local to the surface and the contact pressure generates tensile and compressive stresses that are concentrated near the edge-of-contact between the two surfaces

(Chan et al. 2001). These highly concentrated stresses are in some ways analogous to those that occur in sharply notched members. The local contact stresses in fretting fatigue are sensitive to the geometry of the two surfaces in contact, just as the local notch stresses are sensitive to the details of the notch geometry. As one would expect, these concentrated stresses govern both crack initiation and early crack growth in fretting fatigue. Numerous studies have indicated that these cracks are initiated relatively early in the life of the specimen under both low- and intermediate-cycle fretting fatigue, as well as high-cycle fretting fatigue (Chan et al. 2001; Hudak et al. 2002).

The application of small-crack behavior to fretting fatigue is conceptually similar to the application to notch fatigue. However, the implementation of these concepts is more complex for the case of fretting fatigue due to the added complexities of the mechanics of contact fatigue. The primary additional complexity in fretting fatigue is the existence of two independent stress fields – one due to the contact pressure and one due to the remotely applied fatigue loading. This gives rise to more complex stresses than in the case of fatigue notches. For example, the mean stress or the ratio,  $R$ , changes even for remote constant amplitude fatigue loading and cracks are generally loaded under mixed mode conditions. Nevertheless, these complexities can be treated as described in detail elsewhere (Chan et al. 2001; Hudak et al. 2002). Like FOD, a treatment of small-crack effects and the associated conditions for non-propagation of cracks in the stress gradients of the contact stress field due to fretting had led to the definition of a limiting “worst-case fret” (Chan et al. 2001; Hudak, et al. 2002).



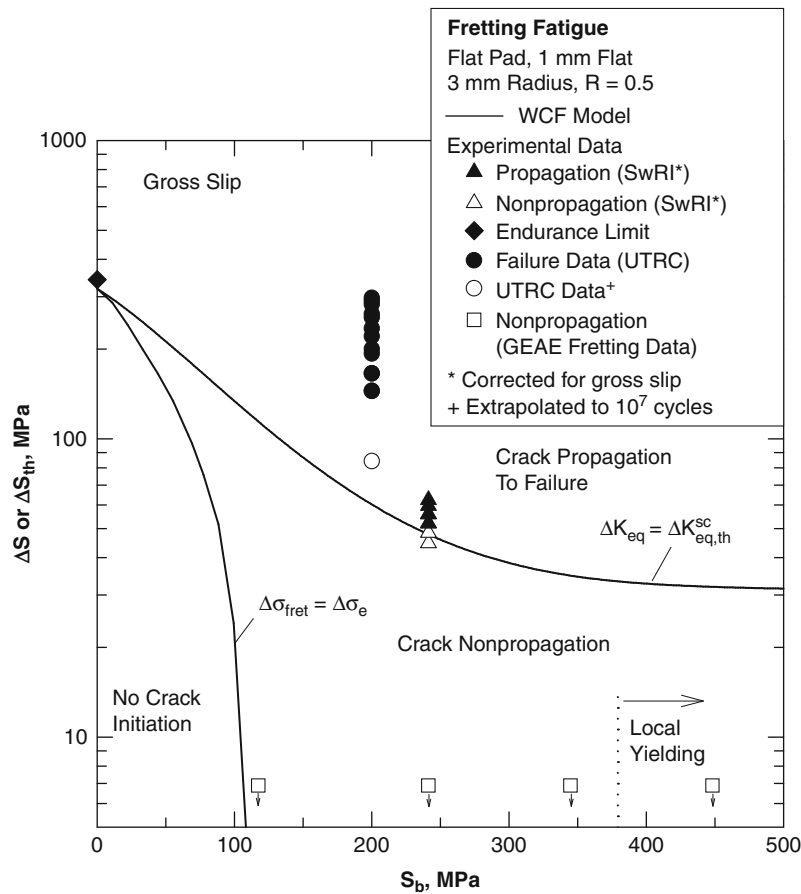
**Growth Characteristics of Small Fatigue Cracks, Fig. 14**  
Calculated threshold stress,  $\Delta S_{th}$ , for crack nonpropagation as a function of crack length for a 1 mm flat pad subjected to various bearing pressures and shear forces (Chan et al. 2001)

The predicted dependence of threshold stress,  $\Delta S_{th}(a)$ , versus crack length curves on the applied bearing stress,  $S_b$ , is shown in Fig. 14. These results were computed for the specific case of a 1.0-mm flat pad with 3.0-mm radii of curvature on each edge, and assumed nominal fatigue properties for Ti-6Al-4V at  $R = 0.5$ . For low to moderate bearing stresses, the  $\Delta S_{th}(a)$  curves decrease with increasing bearing stress and the maximum value of  $\Delta S_{th}$  occurs for vanishingly small crack sizes indicating that the fatigue life is controlled by crack initiation. In contrast, as the bearing stress continues to increase, the  $\Delta S_{th}(a)$  curves exhibit a local maximum at increasing distances below the surface. This behavior indicates that cracks can initiate and arrest at high bearing stresses, and is strikingly similar to the behavior of notches. The dependence of the threshold stress in Fig. 14 on bearing stress is due to the influence of bearing stress on the magnitude of the contact stress field and is analogous to the influence of notch severity on the local stress field and stress gradient of the notch. However, unlike the notch case, the threshold stresses in Fig. 14 approach zero for high bearing stresses and small crack sizes. This behavior indicates that pure fretting (i.e., without the application of a bulk fatigue stress range) is capable of initiating cracks whose equilibrium sizes increases with increasing bearing stress as indicated by the crack sizes at  $\Delta S_{th}(a) = 0$  in Fig. 14.

It is significant to note that as the bearing stress achieves increasingly higher values, the peak in the  $\Delta S_{th}(a)$  curve achieves a limiting value  $\Delta S_{th}^*$  that defines the “worst-case fret” for this particular loading condition and contact geometry. This is more clearly seen in Fig. 15, where the threshold stress is shown as a function of the bearing stress. The curve defining crack initiation from a classical S-N approach ( $\Delta\sigma_{fret} = \Delta\sigma_e$ ) is also shown for comparison with the small crack fracture mechanics approach ( $\Delta K_{eq} = \Delta K_{eq,th}^{SC}$ ). These two curves define three regimes depicting no crack initiation, crack initiation and arrest, and crack propagation to failure. The model predictions in Fig. 15 are compared with fretting fatigue data from the literature. The data are in good agreement with the model predictions. The predicted saturation in the threshold stress for specimen failure with increasing bearing stress is also in agreement with fretting fatigue data on a wide range of materials.

The primary feature that sharp notches and fretting fatigue have in common that makes them amenable to treatment using small-crack fracture mechanics is their high stress concentration and steep stress gradient. The concentrated stress at the surface facilitates early initiation of microcracks, particularly under service loading, which contains low cycle fatigue in addition to high cycle fatigue. At notches, the region of high stress gradient is on the order of the notch root radius. For FOD-induced notches this size scale is on the order of 0.1–1 mm. Fretting fatigue with comparable stress concentrations results in stress gradients that are at least as steep as those at sharp notches. In order to ensure sufficient accuracy, a fracture mechanics treatment of such stress steep gradients should consider cracks that are about 10% of the size scale over which the gradient occurs (0.01–0.1 mm). Thus, any fracture mechanics treatment of these high stress gradients would need to consider small-crack behavior in order to be physically realistic.

The essence of the small-crack fracture mechanics approach to HCF is to differentiate between cracks that will arrest after a small amount of growth versus cracks that will continue to grow to failure. This differs from the classical fracture mechanics approach to damage tolerance in two respects. First, HCF is by its very nature a threshold problem, not a traditional crack growth problem. This is because HCF occurs at cyclic frequencies of 1,000 Hz or more. At these high frequencies, the crack growth rate on a time basis is relatively fast. Thus, once sufficient growth has occurred, failure is eminent. Under these conditions, periodic inspection for HCF cracks is not economically viable since the inspection intervals would be prohibitively short. Second, the crack sizes of relevance to the methods described herein are below the current threshold for



**Growth Characteristics of Small Fatigue Cracks, Fig. 15** Predicted threshold stresses,  $\Delta S_{th}$ , for crack initiation, nonpropagation and propagation to failure as a function of bearing stress,  $S_b$ , compared to fretting fatigue (UTRC and SwRI data) and fretting data (GEAE data). Fretting and fretting fatigue data were generated using a 1 mm flat pad with 3 mm radius rounded edges (Chan et al. 2001)

reliable detection of cracks. Thus, even if the inspection intervals were not prohibitively short, the current inspection sensitivity is still inadequate. On the other hand, HCF generally occurs at the resonant frequencies, but the time a system spends at resonance is usually very short so the total number of high-cycle fatigue cycles encountered during a system resonance event may be small and is less than that required to reach the critical crack length. Under this circumstance, fatigue crack growth under HCF loads can be treated in a similar manner as those employed for treating FCG under LCF loading.

## Cross-References

- [Fatigue](#)
- [Fatigue Limit](#)
- [Growth Characteristics of Large Fatigue Cracks](#)
- [Stress Intensity Factors](#)

## References

- K.S. Chan, M.P. Enright, A probabilistic micromechanical code for predicting fatigue life variability: model development and application. *J. Eng. Gas Turbines Power* **129**, 889 (2006)
- K.S. Chan, J. Lankford, The role of microstructure dissimilitude in fatigue and fracture of small cracks. *Acta Metall.* **36**, 193 (1988)
- K.S. Chan, Y.D. Lee, D.L. Davidson, S.J. Hudak Jr., A fracture mechanics approach to high cycle fretting fatigue based on the worst case fret concept Part I – model development; Part II – experimental evaluation. *Intern. J. Fract.* **112**, 299 (2001). p. 331
- D.L. Davidson, K.S. Chan, S.J. Hudak Jr., R.C. McClung, Small fatigue cracks, in *Comprehensive Structural Integrity*, ed. by R.O. Ritchie, Y. Murakami. Cyclic Loading and Fatigue, vol. 4 (Elsevier, Oxford, 2003), p. 129
- M.H. El Haddad, K.N. Smith, T.H. Topper, Fatigue crack propagation of short cracks. *ASME J. Eng. Mater. Technol.* **101**, 42 (1979)
- R.P. Gangloff, Effect of crack size on the chemical driving force for corrosion fatigue. *Metall. Trans.* **16A**, 953 (1985)
- S.J. Hudak Jr., K.S. Chan, G.G. Chell, Y.-D. Lee, R.C. McClung, A damage tolerance approach for predicting the threshold stresses for high cycle

fatigue in the presence of supplemental damage, in *Proceedings of David L. Davidson Symposium on Fatigue*, ed. by K.S. Chan, P.K. Liaw, R.S. Bellows, T.C. Zogas, W.O. Soboyejo (TMS, Warrendale, 2002), p. 107

S.K. Jha, M.J. Caton, J.M. Larsen, A new paradigm of fatigue variability behavior and implications of life prediction. *Mater. Sci. Eng. A* **468–470**, 23 (2007)

H. Kitagawa, S. Takahashi, in *Proceedings of the 2nd International Conference on Mechanical Behavior of Materials*, Boston, 1967, p. 627

J. Lankford, The growth of small fatigue cracks in 7075-T6 aluminum. *Fatigue Eng. Mater. Struct.* **5**, 233 (1982)

J. Lankford, S.J. Hudak Jr., Relevance of the small crack problem to lifetime prediction in gas turbines. *Int. J. Fatigue* **9**, 87 (1987)

K.J. Miller, Materials science perspective of metal fatigue resistance. *Mater. Sci Technol.* **9**, 453 (1993)

J.C. Newman Jr., Fatigue-life prediction technology using a crack-closure model. *J. Eng. Mater. Technol.* **117**, 433 (1995)

S. Suresh, R.O. Ritchie, The propagation of short fatigue cracks. *Int. Metall. Rev.* **29**, 445 (1984)

---

## GS-Gecko Surface

► [Gecko Toe Surface](#)

---

## GW Theory

► [Contact of Rough Surfaces: The Greenwood and Williamson/Tripp, Fuller and Tabor Theories](#)



